

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 6, 2004, 20:15:58 ; Search time 130 Seconds  
(Without alignments)  
804.498 Million cell updates/sec

Title: US-09-938-901a-2

Perfect score: 1701  
Sequence: 1 MEMRKALLAWYRENNARPLP.....VLRKALLPLAHAGVPLPDA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodaca/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodaca/2/pubpaa/PTC\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodaca/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodaca/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodaca/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodaca/2/pubpaa/PTCUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodaca/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodaca/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodaca/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodaca/2/pubpaa/US09C\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodaca/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodaca/2/pubpaa/US10A\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodaca/2/pubpaa/US10B\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodaca/2/pubpaa/US10C\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodaca/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodaca/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodaca/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodaca/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1697	99.8	325	10	US-09-938-901-2
2	483.5	28.4	486	9	US-09-925-301-1326
3	477	28.0	375	15	US-10-289-762-421
4	448	26.3	482	12	US-10-425-114-68280
5	442	26.0	474	16	US-10-437-963-13871
6	389.5	22.9	188	15	US-10-629-951-36
7	377	22.2	313	14	US-10-156-761-12241
8	365	21.8	293	9	US-09-738-626-6433
9	337	19.5	1719	16	US-10-437-963-103808
10	263.5	15.5	184	12	US-10-424-599-248081
11	240.5	14.1	188	15	US-10-629-951-34
12	234	13.8	230	12	US-10-335-977-7283
13	228.5	13.4	187	12	US-10-335-977-7282
14	200.5	11.8	98	12	US-10-335-977-7279
15	200.5	11.8	98	12	US-10-335-977-7280

16	200.5	11.8	214	12	US-10-335-977-7461	Sequence 7461, Ap
17	199	11.7	268	9	US-09-864-866-43	Sequence 43, Appl
18	191	11.2	185	15	US-10-629-951-35	Sequence 35, Appl
19	184.5	10.8	310	14	US-10-156-761-12127	Sequence 12127, A
20	176.5	10.4	90	11	US-09-864-408A-9042	Sequence 9042, Ap
21	176	10.3	259	15	US-10-369-493-5691	Sequence 5691, Ap
22	163.5	9.6	304	12	US-10-424-599-280749	Sequence 280749, A
23	158.5	9.3	355	15	US-10-369-493-2113	Sequence 2113, Ap
24	155	9.1	172	12	US-10-424-599-150617	Sequence 150617, A
25	155	9.1	393	16	US-10-437-963-113667	Sequence 65596, A
26	150	8.8	373	12	US-10-425-114-65596	Sequence 3828, Ap
27	147.5	8.7	260	9	US-09-738-626-3828	Sequence 296, Appl
28	145	8.5	211	9	US-09-912-020-296	Sequence 246739, A
29	137.5	8.1	184	15	US-10-629-951-33	Sequence 39640, A
30	130	7.6	158	12	US-10-424-599-246739	Sequence 11, Appl
31	129.5	7.6	170	16	US-10-767-701-33640	Sequence 3118, Ap
32	129.5	7.6	5245	14	US-10-329-079-11	Sequence 8128, Ap
33	128	7.5	281	14	US-10-128-714-3128	Sequence 13299, A
34	128	7.5	461	14	US-10-128-714-8128	Sequence 61161, A
35	126	7.4	272	15	US-10-369-493-11299	Sequence 70995, A
36	125.5	7.4	148	16	US-10-767-701-61161	Sequence 68053, A
37	125.5	7.4	6238	16	US-10-343-710-71	Sequence 43512, A
38	120.5	7.1	776	12	US-10-425-114-70995	Sequence 114893, A
39	119	7.0	187	12	US-10-425-114-68053	Sequence 71777, A
40	118.5	7.0	416	12	US-10-425-114-43512	Sequence 11, Appl
41	117.5	6.9	1072	16	US-10-437-963-114893	Sequence 11, Appl
42	112.5	6.6	440	12	US-10-425-114-71777	Sequence 44928, A
43	112.5	6.6	1332	10	US-09-840-743-11	
44	111	6.5	1729	10	US-09-840-743-2	
45	110.5	6.5	457	12	US-10-282-122A-44928	

# ALIGNMENTS

RESULT 1  
US-09-938-901-2  
Sequence 2, Application US/09938901  
Publication No. US20030008291A1  
GENERAL INFORMATION:  
APPLICANT: Yokoyama Shigeyuki  
TITLE OR INVENTION: GENE ENCODING DNA REPAIR ENZYME  
FILE REFERENCE: PH-1261-US  
CURRENT APPLICATION NUMBER: US/09/938, 901  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: JP2001-47762  
PRIOR FILING DATE: 2001-02-23  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 325  
TYPE: PRT  
ORGANISM: Thermus thermophilus  
US-09-938-901-2

Qy	Query Match	99.8%	Score 1697	DB 10	Length 325
Db	Best Local Similarity	99.7%	Pred. No. 2e-151		
	Matches 324	Conservative 1	Mismatches 0	Indels 0	Gaps 0
Qy	1	MEMRKALLAWYRENNARPLPWRGKDYRLVSEVLLQOTRVGOALPYRRFLERFPTLK	60		
Db	1	VEAMRKALLAWYRENNARPLPWRGKDYRLVSEVLLQOTRVGOALPYRRFLERFPTLK	60		
Qy	61	ALAAASLEEVLRWQAGAGYRRAEHLRLARVSEELPPSFARLGLPGLAGPYAAVAASI	120		
Db	61	ALAAASLEEVLRWQAGAGYRRAEHLRLARVSEELPPSFARLGLPGLAGPYAAVAASI	120		
Qy	121	AFGERVAADVGNVRRVLSRLPARESPEKELFALAOGLPEGVDPGVWQNALMELGATVC	180		
Db	121	AFGERVAADVGNVRRVLSRLPARESPEKELFALAOGLPEGVDPGVWQNALMELGATVC	180		
Qy	181	LPRKPRGACPLGAFCKGKAPGRYPAPRRRAKKEELVALVLGRKGVHLERLEGRFOG	240		

```

Db      181 LPRPRGACPLGAFGRGKAPRYPAPRRKAKRELVALVLGKQVHLRLGRFGQ 240
Qy      241 LVGVPLPPEELPGREAAFGVRSRPLGEVHALTHRLRLVEVRGALMEGEGEDPKRPLP 300
Db      241 LVGVPLPPEELPGREAAFGVRSRPLGEVHALTHRLRLVEVRGALMEGEGEDPKRPLP 300
Qy      301 KLMEXKLRKALPLLAAGVPLPDA 325
Db      301 KLMEXKLRKALPLLAAGVPLPDA 325

```

# RESULT 2

```

US-09-925-301-1326
; Sequence 1326, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1999-03-12
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1326
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (438)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (447)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1326

```

Query Match 28.4%; Score 483.5; DB 9; Length 486;  
Best Local Similarity 36.5%; Pred. No. 1e-36;

Matches 135; Conservative 46; Mismatches 106; Indels 83; Gaps 13;

```

Qy      3 AMRKALLAWYRENAARPLPWG-----EKDPYLVSEVVLQOTRVEQALPYRRFLE 54
Db      106 AFRGSLISWYDQKRDLPKRRRAEDMDLDRAYAVVWSEVWLQOTQVATVINYYTGMQ 165
Qy      55 RFPPTLKALAAASLEEVLRWQAGYVRAEHLRLARS-VEEL--PSPFAEL--RGLPG 108
Db      166 KWPPTLDLASSASLEEVNQWLAGLGYYSRGRRLQEGARKVVEELGMPPTAATLQQLPG 225
Qy      109 LGGYTAAVAASIAFGERVAAVGNVRYLSRLFA-----RSPKEXELFALAQGLPEGV 163
Db      226 VGGYTGALIASIAFGATGVVDGNVAVLCRVATIGADPSSSTVLGQQLMGLAQQLV-DPA 284
Qy      164 DPGVNNQALMELGATVCLPRPRPGACPLGAFGRGK-----A 201
Db      285 RPDGFWQAMELGATCTPQRPLCGSCCPVSLCRARQREGRQQLASGLSSPVEECA 344
Qy      202 P-----GRYPAPRRKRAK--EERLVALVL-----LGRKGVHLER- 233
Db      345 PVTGQCHCLPPESEPDQTLGVVNFPRKRSRKPRESSATVLEQPGALGQILVGRP 404
Qy      234 LGRFGQGLYGVV--LPPPEELPGREAF-----GVRSRPLGEVHALTHRLRLV 280
Db      405 NSGLLAGLMEFPVSVWSEBQQRKALLQELQDAXAGPLPATHTXRLRGVHTFSHIKLTY 464
Qy      281 EYRGALMEGE 290

```

```

Db      465 QVYGALMEGQ 474

```

# RESULT 3

```

US-10-289-762-421
; Sequence 421, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 421
; LENGTH: 375
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-289-762-421

```

Query Match 28.0%; Score 477; DB 15; Length 375;  
Best Local Similarity 38.6%; Pred. No. 3e-36;  
Matches 117; Conservative 44; Mismatches 112; Indels 30; Gaps 7;

```

Qy      6 KALLAWYRENAARPLPWGKDPYVIVSEVVLQOTRVEQALPYRRFLEPPTLKALAA 65
Db      22 EALKKPFKRNKRSPLPKRNPTPTVYVWSEVWLQOTRAEVVIDFNMWMEFPPTIESLAA 81
Qy      66 SLEEVLRWQAGYVRAEHLRLARSVEE--LPPSFAELRGLPGLPYTAAVAASI 120
Db      82 KEEDVITKIMEGLGYYSRAPHLLLEGARVWMEHPGKIDDAISLAQRGVGPYVHALAR 141
Qy      121 AFGERVAAVDGNVRYLSRLPARESPEKE-----LFLAQGLPEGVDCVNNQALMEL 175
Db      142 AFRRAAAVDGNVRYLSRLPLETSTIDLESTRTWVSRIAQLLPH-KSPREVIAELIEL 200
Qy      176 GATVCLPRPRGACPLGAFGRGKAPRYPAPRRKRAK-----ERLVALVL--GRKGV 230
Db      201 GACIC-KVFPQCHRCFVROACGAMRENKQFVLVRRARKKVTILHLVALVLYDGLVVE 259
Qy      231 LERLEGRFQGLYGVPLPPEELPGREAAFGVRSR-----PLGEVHALTHRR 277
Db      260 KRRPKENMAGLYERPYLEVSPBEDLQDIEGFTKMELESPLFLGNLKEQGHAFTHNK 319
Qy      278 LRV 280
Db      320 VHL 322

```

# RESULT 4

```

US-10-425-114-68280
; Sequence 68280, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yina
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaka, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 2003-04-28
; SEQ ID NO 68280
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Zea mays

```

```

; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73265D07_FLI.pep
US-10-425-114-68280
Query Match      26.3%; Score 448; DB 12; Length 482;
Best Local Similarity 35.4%; Pred. No. 2,3e-33;
Matches 123; Conservative 44; Mismatches 113; Indels 58; Gaps 10;

QY 3 AMRKALLAWYRENARPLPWR-----GKDPYRVLVSEVLLQOTRVEQALPYRRFLERP 57
DB ALAAQQLRMVDARHRLDPMRCVSGSSEERAYAVWVSEVLLQOTRVPVVAAYERRMARWP 129
QY 58 TLKALAAASLEBVLRWQAGYRRARHLRLARSV---LPPSFELGGLGLGYTA 114
DB TVRSLAAATOEVEVEMWAGYRRARFLLEGAKQIIEKGLPCTALALAEVVGIGDYTA 189
QY 115 AAVASIAFGEVVAADVGNVRVLSRLFA-RESPKEKELFALAQGLPEGV---PGVMNQ 170
DB GALASIAFNEVVPVDGNVIRVLSRLTYTTIADNPKESSTVKRFMDLVQGMVDPLRPGDFNQ 249
QY 171 ALMELGATVCLPKRPGACPLGAF-----RGKAPGRYPAPR---KRAREERLVAL 221
DB 250 AMMELGATLSCKTPGSCQPVSHCOALALSRKESVQYTDPRVVPKAKPRSDFAVC 309
QY 222 VLLGRKV-----HLERL-----ECRFQGLYGVLPFPE---ELPGREAFG- 260
DB 310 VVOQAQGLEBAADPKGNHFLILIKRPEGLAGLWEPVLVVDQKTDLLNRKXMDK 369
QY 261 -----VRSRPLGEVRHALTHRLRVEV 282
DB 370 YLSKLSIDMVRKPDVILREDDVGHVHIFSHIRLTMHV 407

RESULT 5
US-10-437-963-138271
; Sequence 138271, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 138271
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_39677C.1.pap
US-10-437-963-138271
Query Match      26.0%; Score 442; DB 16; Length 474;
Best Local Similarity 35.0%; Pred. No. 8,3e-33;
Matches 122; Conservative 47; Mismatches 110; Indels 70; Gaps 12;

QY 3 AMRKALLAWYRENARPLPWRGEKDP-----YRVLVSEVLLQOTRVEQALPY 48
DB 56 AYRRELRLWYDNRHRLDPMRAAEPPAGSGSGRGEQAYAVWVSEVLLQOTRVPVVDY 115
QY 49 YRRPLERPPTLKALAAASLEBVLRWQAGYRRARHLRLARSV---ELPPSPFALRG 105
DB 116 YSRWMAARPTVDSLAATOEVEVEMWAGYRRARFLLEGAKQIIEKGEFPCTASTLR 175
QY 106 LPGAQPYTAAVAASIAFGEVVAADVGNVRVLSRLFA-RESPKEKELFALAQGLPEGV 164

```

```

DB 176 VARGIGDYTGALASIAFNFVVPVDGNVRVLSRFYALPDNPKESSTVKRFMQLNSELVD 235
QY 165 ---PGVMNQALMELGATVCLPKRPGACPLGAFGRKAPGR-----YR--APRR 211
DB 236 PSRPGDFNQAMMELGATLSCKTPGSCQPVSHCOALALSSQNASVKYTDPRVVPKAK 295
QY 212 RAKEERLVALVYLLGRGVHLERLEGRFQGLYGVLPPEPELPGREAFGRSRPLGEVRH 271
DB 296 PMSDFPAAVCVOIS-----QGRGEGI-----AEAGSKNLFILIKRP----- 332
QY 272 ALTHRLRLEVRGALWE-----GEGE-DPMKRPLPKMEKYLKRALPL 313
DB 333 -----EKGGLAGLWEPFVLVNEKTDTLNR--RKENDKYLKQULSI 372

RESULT 6
US-10-629-951-36
; Sequence 36, Application US/10629951
; Publication No. US20040018550A1
; GENERAL INFORMATION:
; APPLICANT: Bellacosa, Alfonso
; TITLE OF INVENTION: Methods for Detection of Transition
; TITLE OF INVENTION: Single-Nucleotide Polymorphisms
; FILE REFERENCE: FCCC 96-21
; CURRENT APPLICATION NUMBER: US/10/629,951
; CURRENT FILING DATE: 2003-07-29
; PRIOR APPLICATION NUMBER: US/09/629,222A
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/463,891
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US98/15828
; PRIOR FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: 60/053,936
; PRIOR FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 188
; TYPE: PRT
; ORGANISM: E. coli
US-10-629-951-36
Query Match      22.9%; Score 389.5; DB 15; Length 188;
Best Local Similarity 44.8%; Pred. No. 2,2e-28;
Matches 81; Conservative 37; Mismatches 50; Indels 13; Gaps 5;

QY 27 PYRVLVSEVLLQOTRVEQALPYRRFLERPPTLKALAAASLEBVLRWQAGYRRARHL 86
DB 2 PYRVLVSEVLLQOTQVATVPIYFERPMARPTVTDLANAPLDEVLHMTGLGYARARNT 61
QY 87 HRLARVSEEL-----PPSFAELRGLPGLGYTAANAASIAFGEVVAADVGNVRVLSRLF 141
DB 62 HKAAQVATLHGKFPTEFEVVALPGVGRSTAGALISLKGKFPILDNVVKVRLARCY 121
QY 142 ARRS-PKEK-----LFLAAGLPL- GYDGVGNQALMELGATVCLPKRPGACPLGAF 195
DB 122 AVSGWPGKEVENKMTLSLSQVTPAVGVER--FNQAMMDIGAMI CTRSKRKCISLCPLONG 179
QY 196 C 196
DB 180 C 180

RESULT 7
US-10-156-761-12241
; Sequence 12241, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI

```





Db 174 EVRGIDYTAGAIAIAFNEVVPVQGNV-RVLSRLYALPDNKE----- 217  
 QY 164 DPGVNNQALMELGATVCLPFRPGACPLGAFCKRGKAPGRYPAPRRR-----A 213  
 Db 218 -----SSTV---KR-----FCRRQQADELAAAEERRRATARTGMAAA 252  
 QY 214 KEERLYALVLGRKG-----VHLERLEGKQGLYGVPLFPPELPGREAA 258  
 Db 253 RAARLAAAEIAAARAEEAEADAARAEEVEVETLRSSINGSIAGDITADREL--EELA 310  
 QY 259 FGVRSRPLGE 268  
 Db 311 RG-RARERAE 319

## RESULT 10

US-10-424-599-248081  
 ; Sequence 248081, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 248081  
 ; LENGTH: 184  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_66047C.1.pep  
 ; US-10-424-599-248081

Query Match 15.5%; Score 263.5; DB 12; Length 184;  
 Best Local Similarity 45.5%; Pred. No. 1.6e-16;  
 Matches 56; Conservative 19; Mismatches 35; Indels 13; Gaps 2;

QY 5 RKALLARYENARPLPRG-----EKDPYVVLVSEVLQOTRYEOLPYRRRLER 55  
 Db 62 RVALLDWYDINRRDLPRRTTFKQDEEYERRAYGVWVSEVWLQOTRYOTVAYYNNRWOK 121  
 QY 56 PPTLKALAAASLEEVLAVMWGAGYRRARHRLARSV---ELPPSPALRGLPGIGP 111  
 Db 122 WPTIHLLAQASLEEVNEMWAGLYRRRARRFLLEGAKKIVAEGGQIPKVASMLRNIPGIGE 181  
 QY 112 YTA 114  
 Db 182 YTS 184

## RESULT 11

US-10-629-951-34  
 ; Sequence 34, Application US/10629951  
 ; Publication No. US20040018550A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bellacosa, Alfonso  
 ; TITLE OF INVENTION: Methods for Detection of Transition  
 ; TITLE OF INVENTION: Single-Nucleotide Polymorphisms  
 ; FILE REFERENCE: FCCC 96-21  
 ; CURRENT APPLICATION NUMBER: US/10/629,951  
 ; CURRENT FILING DATE: 2003-07-29  
 ; PRIOR APPLICATION NUMBER: US/09/629,222A  
 ; PRIOR FILING DATE: 2000-07-31  
 ; PRIOR APPLICATION NUMBER: 09/463,891  
 ; PRIOR FILING DATE: 2000-01-28  
 ; PRIOR APPLICATION NUMBER: PCT/US98/15828  
 ; PRIOR FILING DATE: 1998-07-28  
 ; PRIOR APPLICATION NUMBER: 60/053,936

; PRIOR FILING DATE: 1997-07-28  
 ; NUMBER OF SEQ ID NOS: 73  
 ; SOFTWARE: FASTSEQ for Windows Version 3.0  
 ; SEQ ID NO 34  
 ; LENGTH: 188  
 ; TYPE: PRT  
 ; ORGANISM: M. thermoformicium  
 ; US-10-629-951-34

Query Match 14.1%; Score 240.5; DB 15; Length 188;  
 Best Local Similarity 32.0%; Pred. No. 2.4e-14;  
 Matches 58; Conservative 35; Mismatches 77; Indels 11; Gaps 4;

QY 26 DPTRVVSEVLQOTRYEOLPYRRRLERFPTLKALAAASLEEVLAVMWGAGY-YRRAR 84  
 Db 1 DPTVILITELILRRITAGHVKKIYDFEVKTCFEDILTPKSEIKDKIETGLSNORAB 60  
 QY 85 HLHRLARSV-----ELPPSPALRGLPGIGPYTAAVAASIAFGEVAAVQGNVRVLSR 139  
 Db 61 QLKELARVYINDYGVPPNRRKAIIDLPGVKYKTCAAVWGLAFGKKAAMVDANFVAVINR 120  
 QY 140 LFARE----SPKRELPAALQGLLPFGVDPGVNNQALMELGATVCLPKRPGACPLGAF 195  
 Db 121 YFGSYENINYNHKAALMELAEITLVPQKCRD-FNLGLMDFSAITICAPRRKCKEKGMSKL 179  
 QY 196 C 196  
 Db 180 C 180

## RESULT 12

US-10-335-977-7283  
 ; Sequence 7283, Application US/10335977  
 ; Publication No. US20040052799A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DOUGLAS SMITH et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
 ; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR  
 ; DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: LAHIVE & COCKFIELD  
 STREET: 28 State Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02109-1875

COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: IBM PC Compatible  
 OPERATING SYSTEM: Windows NT 4.0  
 SOFTWARE: UNIX

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/335,977

FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:  
 NAME: Mandragouras, Amy E.

REFERENCE/DOCKET NUMBER: 36,207

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)227-7400

TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 7283:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 230 amino acids

TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...230  
SEQUENCE DESCRIPTION: SEQ ID NO: 7283;  
US-10-335-977-7283

Query Match 13.8%; Score 234; DB 12; Length 230;  
Best Local Similarity 30.6%; Pred. No. 1.3e-13;  
Matches 66; Conservative 32; Mismatches 54; Indels 64; Gaps 9;

QY 95 ELPPSPFAELGRLPGCPYTAATAVAASIAFGERYAAVDGVRVLSRLFARSPKKEKLFAL 154  
DB 7 QLPNDYQSLKLPGLGAYTANAILCFGRKSAACVDANVKVLRPF----- 53  
QY 155 AAGLLPBGVDPGV-----WQALMELGATVCLPKRPGACPLGA 194  
DB 54 -----GLDPNTHAKDQIKANDPLNSESFNHQAIDIGALICSPK-PKCAICPFNP 105  
QY 195 FCRGKEAPGRYPAPKRK-AKEERLVALVLGRGVHLERLEGRFOGLY-GVPLFP----- 248  
DB 106 YCLGNHLEHRTLKKKQEIIOERYLGVIONNO-IALEKIE---QKLYIGMHHPNKE 161  
QY 249 --PEELPGRAPAFGRSRPLGEVHALTHRLAVEV 282  
DB 162 NLEFKLPF-----LGTTHSHITKFLNLNL 186

## RESULT 13

US-10-335-977-7282  
Sequence 7282, Application US/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy B.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)742-4214  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 7282;  
SEQUENCE CHARACTERISTICS:  
LENGTH: 187 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:

NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...187  
SEQUENCE DESCRIPTION: SEQ ID NO: 7282;  
US-10-335-977-7282

Query Match 13.4%; Score 228.5; DB 12; Length 187;  
Best Local Similarity 33.5%; Pred. No. 3.3e-13;  
Matches 59; Conservative 25; Mismatches 45; Indels 47; Gaps 7;

QY 95 ELPPSPFAELGRLPGCPYTAATAVAASIAFGERYAAVDGVRVLSRLFARSPKKEKLFAL 154  
DB 16 QLPNDYQSLKLPGLGAYTANAILCFGRKSAACVDANVKVLRPF----- 62  
QY 155 AAGLLPBGVDPGV-----WQALMELGATVCLPKRPGACPLGA 194  
DB 63 -----GLDPNTHAKDQIKANDPLNSESFNHQAIDIGALICSPK-PKCAICPFNP 114  
QY 195 FCRGKEAPGRYPAPKRK-AKEERLVALVLGRGVHLERLEGRFOGLY-GVPLFP 248  
DB 115 YCLGNHLEHRTLKKKQEIIOERYLGVIONNO-IALEKIE---QKLYIGMHHPNKE 166

## RESULT 14

US-10-335-977-7279  
Sequence 7279, Application US/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy B.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)742-4214  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 7279;  
SEQUENCE CHARACTERISTICS:  
LENGTH: 98 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...98  
SEQUENCE DESCRIPTION: SEQ ID NO: 7279;  
US-10-335-977-7279

Query Match 11.8%; Score 200.5; DB 12; Length 98;



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2004, 20:11:37 ; Search time 19 Seconds  
(without alignments)  
883.076 Million cell updates/sec

Title: US-09-938-901a-2

Sequence: 1 MEAMRKALLAWYRENRPLP.....VLRKALPLAHGVPLPDA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	505	29.7	403	4 US-09-107-532A-7308	Sequence 7308, Ap
2	485	28.5	395	4 US-09-134-000C-5115	Sequence 5115, Ap
3	481.5	28.3	535	3 US-08-813-574-2	Sequence 2, Appl1
4	477.5	28.1	470	4 US-09-252-991A-23310	Sequence 23310, A
5	477	28.0	375	4 US-09-198-452A-421	Sequence 421, App
6	471.5	27.7	350	3 US-08-813-574-9	Sequence 9, Appl1
7	471.5	27.7	350	4 US-09-651-656-37	Sequence 37, Appl1
8	471.5	27.7	350	4 US-09-650-855-37	Sequence 37, Appl1
9	457.5	26.9	516	4 US-09-489-039A-11959	Sequence 11959, A
10	450	26.5	347	4 US-09-543-681A-5315	Sequence 5315, Ap
11	443	26.0	346	4 US-09-134-001C-4684	Sequence 4684, Ap
12	442	26.0	346	4 US-09-328-352-4982	Sequence 4982, Ap
13	389.5	22.9	414	4 US-09-629-222A-36	Sequence 36, Appl1
14	355	20.9	414	4 US-09-540-236-2578	Sequence 2578, Ap
15	274.5	16.1	221	1 US-08-663-023-17	Sequence 17, Appl1
16	274.5	16.1	221	1 US-09-402-959A-10	Sequence 10, Appl1
17	274.5	16.1	225	4 US-10-037-927B-14	Sequence 14, Appl1
18	240.5	14.1	188	4 US-09-629-222A-34	Sequence 34, Appl1
19	191	11.2	185	4 US-09-629-222A-35	Sequence 35, Appl1
20	188	11.1	227	4 US-09-107-532A-3843	Sequence 3843, Ap
21	176	10.3	207	2 US-08-808-550-36	Sequence 36, Appl1
22	176	10.3	259	2 US-08-808-550-39	Sequence 39, Appl1
23	172	10.1	209	2 US-08-808-550-31	Sequence 31, Appl1
24	171	10.1	204	2 US-08-808-550-32	Sequence 32, Appl1
25	164.5	9.7	211	2 US-08-808-550-37	Sequence 37, Appl1
26	164.5	9.7	297	2 US-08-808-550-42	Sequence 42, Appl1
27	164.5	9.7	304	2 US-08-808-550-2	Sequence 2, Appl1

28	160	9.4	224	4 US-09-134-001C-3225	Sequence 3225, Ap
29	158	9.3	185	4 US-09-134-000C-3440	Sequence 3440, Ap
30	154	9.1	213	2 US-08-808-550-35	Sequence 35, Appl1
31	145	8.5	207	2 US-08-808-550-29	Sequence 29, Appl1
32	145	8.5	211	2 US-08-808-550-38	Sequence 38, Appl1
33	145	8.5	211	4 US-09-651-656-25	Sequence 25, Appl1
34	145	8.5	211	4 US-09-650-855-25	Sequence 25, Appl1
35	142	8.3	216	4 US-09-489-039A-9560	Sequence 9560, Ap
36	137.5	8.1	184	4 US-09-629-222A-33	Sequence 33, Appl1
37	134	7.9	207	2 US-08-808-550-30	Sequence 30, Appl1
38	128	7.5	628	4 US-09-252-991A-30904	Sequence 30904, A
39	127.5	7.5	212	4 US-09-252-991A-17572	Sequence 17572, A
40	121	7.1	216	4 US-09-543-681A-4806	Sequence 4806, Ap
41	121	7.1	4472	2 US-08-804-227C-2	Sequence 2, Appl1
42	118.5	7.0	776	4 US-09-252-991A-28446	Sequence 28446, A
43	117	6.9	230	4 US-09-328-352-5770	Sequence 5770, Ap
44	115.5	6.8	494	4 US-09-252-991A-17295	Sequence 17295, A
45	115	6.8	542	4 US-09-252-991A-21753	Sequence 21753, A

## ALIGNMENTS

RESULT 1  
US-09-107-532A-7308  
Sequence 7308, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneka  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-5077  
INFORMATION FOR SEQ ID NO: 7308:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 403 amino acids  
TYPE: amino acid  
MOLECULE TYPE: linear  
HYPOTHEICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) Locus 1...403  
SEQUENCE DESCRIPTION: SEQ ID NO: 7308:  
US-09-107-532A-7308



Db 260 REGDFNOAMELGATVCTPDRPLCSQCPVESLCRAORVEQOELLASGLSGSDVEECA 319  
 QY 202 P-----GRYPARKRKRAK---ERLVALVL-----LGRKGVHLER- 233  
 Db 320 PNTGQCHLCIPSEPDQTLGVNFRPKSRKPRESSATCTLEQPGALGAQILLVQRP 379  
 QY 234 LSGRFQGLYGV-LLPPEELPGREAF-----GVSRPLGGEVRAHLTHRLRV 280  
 Db 380 NSGLLGLWEPSPVTPWSEPOLQRKALLOELQWAGPLPATHLRHLEVVHTFSHAKLTY 439  
 QY 281 EVRGALMEGE 250  
 Db 440 QVYGLALEGQ 449

RESULT 4  
 US-09-252-991A-23310  
 ; Sequence 23310; Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 23310  
 ; LENGTH: 470  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-23310

Query Match 28.1%; Score 477.5; DB 4; Length 470;  
 Best Local Similarity 36.2%; Pred. No. 1.3e-39;  
 Matches 126; Conservative 58; Mismatches 119; Indels 45; Gaps 14;

QY 2 EAMRKALLAWY-KENAPPLPWGKEDPYRVLVSEVLLQOTRVEQALPYRRERFPTLK 60  
 Db 119 EGFNGAVLDWYDRGRDXLPWQGITPRVWVSEIMLQOTGVSTVLGPFDFMALPDVE 178  
 QY 61 ALAASLEEVLRVWQAGYVRAHRLRLA-----RSVEELPSPFAELRGPGGYTAA 115  
 Db 179 ALAASLEEVLRVWQAGYVRAHRLRLA-----RSVEELPSPFAELRGPGGYTAA 115  
 QY 116 AVASIAFGERVAADVGNVRVLSRLFARE---SPK-EKELFALAQGLPBGVDPGVWNO 170  
 Db 239 AIALSLNGLRAPILIDGNVRVLSRLFARE---SPK-EKELFALAQGLPBGVDPGVWNO 170  
 QY 171 ALMELGATVCLPKPRCGACPLGAFPCR---GKEAPGRYPAPRKRA-KERLVALVLIG 225  
 Db 298 AMMDLGATLCTRSKPSCLCELVGCRALHGREAD--YPOPKRKALPQKRTIMPILAN 355  
 QY 226 RKG---VHLERLEGRFOGLYGVPLFPPEELPGREAFGRVRSPGGEV-----HALTHR 276  
 Db 356 RDGAILLYRRPSSGGLWGLMSPBL--DDLGLLEPLAARHSLALGERRELSGLTHTSHF 413  
 QY 277 RLRYVE-----VRGA-----LWEGGEDPWKRPLPKMEKYLKRA 310  
 Db 414 QLAIEPWLVAVEGAPRAVABGDWMLVYNLATPP-RLGLAAPVKKLLKRA 460

RESULT 5  
 US-09-198-452A-421  
 ; Sequence 421; Application US/09198452A  
 ; Patent No. 6559294  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Griffiths, R.  
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
 ; TITLE OF INVENTION: and treatment of infection  
 ; FILE REFERENCE: 9710-003-999  
 ; CURRENT APPLICATION NUMBER: US/09/198,452A  
 ; CURRENT FILING DATE: 1998-11-24  
 ; NUMBER OF SEQ ID NOS: 6849  
 ; SEQ ID NO 421  
 ; LENGTH: 375  
 ; TYPE: PRT  
 ; ORGANISM: Chlamydia pneumoniae  
 ; US-09-198-452A-421

Query Match 28.0%; Score 477; DB 4; Length 375;  
 Best Local Similarity 38.6%; Pred. No. 1.1e-39;  
 Matches 117; Conservative 44; Mismatches 112; Indels 30; Gaps 7;

QY 6 KALLAWYRENAPPLPWGKEDPYRVLVSEVLLQOTRVEQALPYRRERFPTLKALAA 65  
 Db 22 BALKKWFENKXKSLPDRNDPTFYSVWVSEVWLQOTRAEVVIDYFNQWMEFPPIESLAA 81  
 QY 66 SLEEVLRVWQAGYVRAHRLRLASVEE-----LPSFAELRGPGGYTAAVASI 120  
 Db 82 KEEDVIKLMWEGLYYSRAHLLGARMVEEFHGKIPDAISLAQIRGVGPTVHALIAP 141  
 QY 121 AGERVAADVGNVRVLSRLFARESPEKE-----LFALAQGLPBGVDPGVWNOALMEI 175  
 Db 142 AFRRAAADVGNVRVLSRLFLIETSIDESTRTVSRINQALLPH-KSEVIAEALIEL 200  
 QY 176 GATVCLPKPRCGACPLGAFPCRGEAPGRYPAPRKRAK---ERLVALVL-GRKGVH 230  
 Db 201 GACIC-KKVPQCHRCVVRQACGAMRENKQFVLRVARKKVIPLHRLVALVYDGSUVVE 259  
 QY 231 LERLEGRFOGLYGVPLFPPEELPGREAFGVNSR-----PLGEVHALTHRR 277  
 Db 260 KRRPKEMAGLYEFPYIEVEPEBGLDIDGFTKMEI-SLESPLEFLGNLKEQSHAFTHNK 319  
 QY 278 LRV 280  
 Db 320 VHL 322

RESULT 6  
 US-08-813-574-9  
 ; Sequence 9; Application US/08813574  
 ; Patent No. 6013473  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wei, Ying-Fei  
 ; TITLE OF INVENTION: Human Mucy  
 ; NUMBER OF SEQUENCES: 18  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SmithKline Beecham Corporation  
 ; STREET: 709 Swedeland Road  
 ; CITY: King of Prussia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19406  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/813,574  
 ; FILING DATE:  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/013,132  
 ; FILING DATE: 11-MAR-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Han, William T  
 ; REGISTRATION NUMBER: 34,344  
 ; REFERENCE/DOCKET NUMBER: ATG50002  
 ; TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5219  
TELEFAX: 610-270-4026  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 350 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-813-574-9

Query Match 27.7%; Score 471.5; DB 3; Length 350;  
Best Local Similarity 35.1%; Pred. No. 3.5e-39;

Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAATRENNAR-PLPWRGKEDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66  
DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVWLQOTVAVIIPYFERFMARFPTVIDLANAP 69  
QY 67 LEEVLRVWQAGYRRAEHLRLARSVEEL-----PSPFAELRGLPGIGPYTAAVASIA 121  
DB 70 LDEVLHLMTGLGYTAAARNLHRAAQOAVATLHGCKPEPTEFEVVALPGVGRSTAGAILSL 129  
QY 122 FGERVAADVGNVRRVLSRLFARSS-PKEKE-----LFALAQGLLPE-GVDPGVNQAALMEL 175  
DB 130 LGKHFPILDGNVRYLARCVAVSGWPGKKEVENKLMISLSEQVTPAVGVER--FNQAMMDL 187  
QY 176 GATVCLPKRPGCAAPL--GAFCRGKEAPGRYPAPRRKRAEERLVALVLLGRGVHL-- 231  
DB 188 GAMICTRSKPKCSLCPLONGCTAANNNSWALYPGKKPKQTLPERTGYFLLQHEDEVLLA 247  
QY 232 -ERLEGRFQGLYGVLPFPEE-----LPGREAFGVSRPLGEVRHALTTRRLRV----- 280  
DB 248 QRPSPGLMGGLYCPQFADBSLRQWLAQRQIADNLTQ-LTAFRHTFSHFHLDIVPMWL 306  
QY 281 -----EVRGALMBEGSDP---WKRPLPKLMEKYLKALPL 313  
DB 307 PVSSFTGCMDEGNALMYNLAQPSVGLAAPVERLLQO-LRTGAPV 350

RESULT 7

US-09-651-656-37  
Sequence 37, Application US/09651656  
Patent No. 6340566  
GENERAL INFORMATION:  
APPLICANT: MCUTHER-MALONEY, SANDRA  
TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE  
TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,  
FILE REFERENCE: IL-10689  
CURRENT APPLICATION NUMBER: US/09/651,656  
PRIOR FILING DATE: 2000-08-29  
PRIOR APPLICATION NUMBER: 60/192,764  
PRIOR FILING DATE: 2000-03-28  
NUMBER OF SEQ ID NOS: 106  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 37  
LENGTH: 350  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-651-656-37

Query Match 27.7%; Score 471.5; DB 4; Length 350;  
Best Local Similarity 35.1%; Pred. No. 3.5e-39;

Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAATRENNAR-PLPWRGKEDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66  
DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVWLQOTVAVIIPYFERFMARFPTVIDLANAP 69  
QY 67 LEEVLRVWQAGYRRAEHLRLARSVEEL-----PSPFAELRGLPGIGPYTAAVASIA 121

DB 70 LDEVLHLMTGLGYTAAARNLHRAAQOAVATLHGCKPEPTEFEVVALPGVGRSTAGAILSL 129  
QY 122 FGERVAADVGNVRRVLSRLFARSS-PKEKE-----LFALAQGLLPE-GVDPGVNQAALMEL 175  
DB 130 LGKHFPILDGNVRYLARCVAVSGWPGKKEVENKLMISLSEQVTPAVGVER--FNQAMMDL 187  
QY 176 GATVCLPKRPGCAAPL--GAFCRGKEAPGRYPAPRRKRAEERLVALVLLGRGVHL-- 231  
DB 188 GAMICTRSKPKCSLCPLONGCTAANNNSWALYPGKKPKQTLPERTGYFLLQHEDEVLLA 247  
QY 232 -ERLEGRFQGLYGVLPFPEE-----LPGREAFGVSRPLGEVRHALTTRRLRV----- 280  
DB 248 QRPSPGLMGGLYCPQFADBSLRQWLAQRQIADNLTQ-LTAFRHTFSHFHLDIVPMWL 306  
QY 281 -----EVRGALMBEGSDP---WKRPLPKLMEKYLKALPL 313  
DB 307 PVSSFTGCMDEGNALMYNLAQPSVGLAAPVERLLQO-LRTGAPV 350

RESULT 8

US-09-650-855-37  
Sequence 37, Application US/09650855  
Patent No. 6363555  
GENERAL INFORMATION:  
APPLICANT: MCUTHER-MALONEY, SANDRA  
TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA  
TITLE OF INVENTION: MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA  
FILE REFERENCE: IL-10284  
CURRENT APPLICATION NUMBER: US/09/650,855  
PRIOR FILING DATE: 2000-08-29  
PRIOR APPLICATION NUMBER: 60/192,764  
NUMBER OF SEQ ID NOS: 106  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 37  
LENGTH: 350  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-650-855-37

Query Match 27.7%; Score 471.5; DB 4; Length 350;  
Best Local Similarity 35.1%; Pred. No. 3.5e-39;

Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAATRENNAR-PLPWRGKEDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66  
DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVWLQOTVAVIIPYFERFMARFPTVIDLANAP 69  
QY 67 LEEVLRVWQAGYRRAEHLRLARSVEEL-----PSPFAELRGLPGIGPYTAAVASIA 121  
DB 70 LDEVLHLMTGLGYTAAARNLHRAAQOAVATLHGCKPEPTEFEVVALPGVGRSTAGAILSL 129  
QY 122 FGERVAADVGNVRRVLSRLFARSS-PKEKE-----LFALAQGLLPE-GVDPGVNQAALMEL 175  
DB 130 LGKHFPILDGNVRYLARCVAVSGWPGKKEVENKLMISLSEQVTPAVGVER--FNQAMMDL 187  
QY 176 GATVCLPKRPGCAAPL--GAFCRGKEAPGRYPAPRRKRAEERLVALVLLGRGVHL-- 231  
DB 188 GAMICTRSKPKCSLCPLONGCTAANNNSWALYPGKKPKQTLPERTGYFLLQHEDEVLLA 247  
QY 232 -ERLEGRFQGLYGVLPFPEE-----LPGREAFGVSRPLGEVRHALTTRRLRV----- 280  
DB 248 QRPSPGLMGGLYCPQFADBSLRQWLAQRQIADNLTQ-LTAFRHTFSHFHLDIVPMWL 306  
QY 281 -----EVRGALMBEGSDP---WKRPLPKLMEKYLKALPL 313  
DB 307 PVSSFTGCMDEGNALMYNLAQPSVGLAAPVERLLQO-LRTGAPV 350

RESULT 9



US-09-489-039A-11959  
 : Sequence 11959, Application US/09489039A  
 : Patent No. 6610836  
 : GENERAL INFORMATION:  
 : APPLICANT: Gary Bretton et al  
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 : TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 : FILE REFERENCE: 2709..2004001  
 : CURRENT APPLICATION NUMBER: US/09/489, 039A  
 : CURRENT FILING DATE: 2000-01-27  
 : PRIOR APPLICATION NUMBER: US 60/117,747  
 : PRIOR FILING DATE: 1999-01-29  
 : NUMBER OF SEQ ID NOS: 14342  
 : SEQ ID NO 11959  
 : LENGTH: 516  
 : TYPE: PRT  
 : ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-11959

Query Match	26.9%;	Score 457.5;	DB 4;	Length 516;
Best Local Similarity	34.8%;	Pred. No. 1.5e-37;		
Matches 120;	Conservative 63;	Mismatches 119;	Indels 43;	Gaps 14;

QY	8	LLAAWRKRNAR--ELPRKGEKDPYRVLYSEVTLLOOTRVEQALPYRRRLKEEFPITLKALAAAS	66
Db	176	VLDMDYDKGRKTLPMQIAKTPPKVWLSVEMLQOOTOVTYIPIFERMANFPIVVDLANAP	235
QY	67	LEEVYRVWQAGAGYVYRRAEHLHRLARSVE-----ELPPFAELRGHFGLGIPYTAANAASIA	121
Db	236	LDEVLHLTGLGLOYARARVHLKAAQOVALTLHGGEPFRITDEVALAPGVGRSTAGIALISLS	295
QY	122	FGEYVAADVGNVRYLSRLFA-----RESPKKELEFALAQGLLP--EGVDPGVWNOALMEL	175
Db	296	LGQHPIILDGNVRCVILARCVAUASGVMPGKKEVEKRLMDISEEVTPAEGVER--FNQAMMDL	353
QY	176	GATVCLPKRPFRGACPL--GAFCRGKEABGRYPAPKPKAKKEBRLVALVLLGRKG--VHL	231
Db	354	GAMVCTRSKPKCELCPLSNGCAVAAVNHSAEYFGKKPKOTLPER--GUYFLIMOHGDEVEL	412
QY	232	EKYLE--GRQGIYGVLPFRPE--ELPRKEAFAFGRSRPLGEV---RHALTHRLRV-----	280
Db	413	SQRPPVGLMGGLFCFPQGFADAELEBRLMIAQROIKANLTLQLTAFRTTFHFHDIIVPMWL	472
QY	281	-----EVRGALMEGEGEDP--WKQRLPYLMEKVRKALPL	313
Db	473	TVHSSGAYMDEGNALMTNLTADQPSVGLAARVERLLQD--LKAGAPV	516

```

RESULT 10
US-09-543-681A-5315
; Sequence 5315, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709, 1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5315
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5315

```

Query Match	26.5%	Score 450;	DB 4;	Length 347;
Best local Similarity	33.2%;	Pred. No. 56-30/;		
Matches 115;	Conservative 64;	Mismatches 123;	Indels 44;	Gaps 12.
QY	2	EAARKALLMAYRENAR-PLPWKGEKDPYRVUJSEVLLQOTRVEQALPYRRFLRSPFLTK	60	

2 EAWRKALLAWYRENAR - PLPWGKDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLK 60

[illegible]

```

RESULT 11
US-09-134-001C-4684
; Sequence 4684, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4684
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4684

Query Match      26.0%; Score 443; DB 4; Length 360;
Best Local Similarity 34.3%; Pred. No. 2,7e-36;
Matches 103; Conservative 56; Mismatches 119; Indels 22; Gaps 7;

```

[illegible]

RESULT 12  
US-09-328-352-4982

```

; Sequence 4982, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAYMANNI FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4982
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4982

```

Query Match 26.0%; Score 442; DB 4; Length 346;  
 Best Local Similarity 33.3%; Pred. No. 3.2e-36;  
 Matches 113; Conservative 60; Mismatches 118; Indels 48; Gaps 11;

```

QY 7 ALAATRENAR-PLPWRGKDPYRLVSEVLQOTRVEQALPYRRFLERPTLKALAA 65
D 11 ALNMPDQGRHDLPMQVADDPYKVVSEIMLQOTVKTLOYFRMEREPTEALGYA 70
QY 66 SIEEVLRVWQAGYVYARAHRLRLASVE--LPPSFAELRGLPGLPYTAAVASIAF 122
D 71 TDEVAIPYAGLGTYARARNLHKAAGLVAAQOKFPETLEEMIALPGIGRTAGALMSLGL 130
QY 123 GERVAADVGNVRVLSRLFARES-----PKEKELFALAGLPGVDGVGNQALMELGA 177
D 131 ROYGVIMGNVGRKVLARFFALIEDLSKQHEREMKLAELCTHRNND--TTOAIMDLGA 189
QY 178 TWCLPRPRGCGACRYGAFSC---RCKEAPRGYRPAKRRAKEEVLVALVLLGRGVHLER 233
D 190 TICTPKKPLCLYCPYQAHQCAVYQGLEDELPKPKTKTPYKTVADVLIIQCEDEMPWQOR 249
QY 234 -LEGFFOGLYGVLPFPE-ELPGREAAFGVSRP-LGEVNHALLH----- 275
D 250 QAHGIMGGLFCLFILENEHERLKUSQCFKQOPQOTQISHFTWLNIAHFVVEBD 309
QY 276 --RLRVNVRGALMEGEGEDPWKRPRLPKLMEKVLKALP 312
D 310 QKEHLAIEBS-----QWLSP-----EOAIKGV 334

```

```

RESULT 13
US-09-629-222A-36
; Sequence 36, Application US/09629222A
; Patent No. 6599700
; GENERAL INFORMATION:
; APPLICANT: Bellacosa, Alfonso
; TITLE OF INVENTION: Methods for Detection of Transition
; FILE REFERENCE: Single-Nucleotide Polymorphisms
; CURRENT APPLICATION NUMBER: US/09/629,222A
; CURRENT FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/463,891
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US98/15828
; PRIOR FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: 60/053,936
; PRIOR FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 188
; TYPE: PRT
; ORGANISM: E. coli
US-09-629-222A-36

```

Query Match 22.9%; Score 389.5; DB 4; Length 188;  
 Best Local Similarity 44.8%; Pred. No. 2.6e-31;  
 Matches 81; Conservative 37; Mismatches 50; Indels 13; Gaps 5;

```

QY 27 PYRLVSEVLQOTRVEQALPYRRFLERPTLKALAAASLEEVLRVWQAGYVYARAH 86
D 2 PYKVMSEVWLQOTVATVPYFPERPMARFPVTDLANPLDEVLHMTGLGYARAKYL 61
QY 87 HRLARSVEE-----PPSFAELRGLPGLPYTAAVASIAFGERVAADVGNVRVLSRLP 141
D 62 HKAAGVATLHGKGFETEBEVAALPVGRTAGAILSLSGKHPPIIDGVNKRVLARCY 121
QY 142 ARS-PKEXE---LFAAGLLE-GYDPGVNQALMELGATVCLPFRPGACPLGAF 195
D 122 AVSGWPKGEVENKMLSLSEQVTPAVGYER--FNQAMMDLGMICTRSPKSCSLCPLONG 179
QY 196 C 196
D 180 C 180

```

```

RESULT 14
US-09-540-236-2578
; Sequence 2578, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2578
; LENGTH: 414
; TYPE: PRT
; ORGANISM: M. catarrhalis
US-09-540-236-2578

```

Query Match 20.9%; Score 355; DB 4; Length 414;  
 Best Local Similarity 28.5%; Pred. No. 2.3e-27;  
 Matches 109; Conservative 57; Mismatches 132; Indels 84; Gaps 13;

```

QY 3 AMRKALAMYRENAR-PLPWRGKDP---YRLVSEVLQOTRVEQALPYRRFLERPP 57
D 33 SFARILLTWPELHGRHGLPMQYHNPASADIVANVSEIMLQOTVYVTLKFEPEPLARFA 92
QY 58 TLKALAAASLEEVLRVWQAGYVYARAHRLRLASVE-----LPPSFAELRGLPGLG 110
D 93 TVGELAVADWQEVASFAGLGYARARNLHGAQGVADFDTHGRFPETVEMQAVKGVG 152
QY 111 PYTAAVASIAFGERVAADVGNVRVLSRLFA-----RESPKEKELFALAGLPEGVDP 165
D 153 RSTAGAILVMGVYKFGVICDGVNKRVLARHRAVCGDITKSAITDKRLMELATALTPEKYS- 211
QY 166 GVNQALMELGATVCLPFRPRGACPLGAFCSGKEAPRGYRPAKRRAKEE---RLVALV 222
D 212 GHYAQMMDLGAITCTRTQPKCHLCPTVDDCAVALGVQSLPVYKKAAPRPYHNSIALS 271
QY 223 LIG---RKGVHLERLEGRFGYGVLPFPEELP-----GREAA 288
D 272 LTHGGLTLHNRQNGGGLWDLGMSLPIF---MLPLDNDOKLDNNTLSDAIKAWQSDKV 348
QY 259 FGVRSRPGEV-----RHALLH-----RLRVEVGA----- 285
D 329 HDLHLSQILELPIPTQTLTAVALRHHTLVHMLYGMSCIANNSQFNQINGTLGLGIDY 388
QY 286 LMEGEGEDPWKRPRLPKLMEKVL 307
D 389 LWT---DTPVNLPLPAMMKLL 407

```

```

RESULT 15
US-08-663-023-17
; Sequence 17, Application US/08663023
; Patent No. 5763178

```

## GENERAL INFORMATION:

APPLICANT: CHIRIKIAN, Jack G.  
APPLICANT: COLLIER, G. Bruce  
TITLE OF INVENTION: OSCILLATING SIGNAL AMPLIFIER FOR NUCLEIC  
TITLE OF INVENTION: ACID DETECTION  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/663,023  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/012,950  
FILING DATE: 06-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/483,089  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 66669/110  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-663-023-17

Query Match 16.1%; Score 274.5; DB 1; Length 221;

Best Local Similarity 31.7%; Pred. No. 1.2e-19;

Matches 63; Conservative 40; Mismatches 85; Indels 11; Gaps 4;

QY 8 LLAWYRENARPLPWRGSKDPYRVIVSEVLLQOTRVEQALPYRRFLERFPTLKALAAASL 67  
DB 16 ILTFMNTDRDPFRHTRDPYVILITELRLRTTAGHVKKIYDKFVVKCFEDILKTPX 75  
QY 68 BEVLRWQAGY-YRRAEHLRLARSV-----ELPPSFAELRGLPGLGPTAAVASIA 121  
DB 76 SEIAKDKEIGLSNQRAEQLELARVINDYGRVPRNRKAILDLPGVGYTCAAVMCLA 135  
QY 122 FGERVAVDGVRVRLSRLFARF-----SPKEKELFALAQGLPBGVDPGVVNOALMELGA 177  
DB 136 FGKAAVVDANFVIVIRYFSGSYENANVNHKALMELAEITVPGKCRD--FNLGLMPSA 194  
QY 178 TVCLPKRRCGACPLGAF 196  
DB 195 IICAPRKKECKGMSKLC 213

Search completed: October 6, 2004, 20:19:07  
Job time : 21 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2004, 04:57:19 ; Search time 564 Seconds  
(without alignments)  
8763.879 Million cell updates/sec

Title: US-09-938-901A-1

Perfect score: 975  
Sequence: 1 atgagagcctgcgcgaagc.....tagtccccctccgcgacgca 975

Scoring table:  
IDENTITY NUC  
Gapop 10\_0, Gapext 1.0

Searched: 3340653 seqs, 2534783454 residues

Total number of hits satisfying chosen parameters: 6681306

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

Published Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	974	99.9	975	US-09-938-901-1	Sequence 4577, App
2	112.4	11.5	939	US-10-156-761-1	Sequence 1, Appl
3	112.4	11.5	9025608	US-10-156-761-1	Sequence 1588, App
4	104	10.7	1854	US-10-469-992-9	Sequence 3596, App
5	104	10.7	1878	US-09-925-301-484	Sequence 3629, App
6	92.8	9.5	528	US-09-974-300-1583	Sequence 42, Appl
7	91.2	9.4	5850	US-10-437-963-1325	Sequence 5404, App
8	89.2	9.1	1660	US-10-425-114-3212	Sequence 1, Appl
9	86.6	8.9	912	US-10-216-817-21	Sequence 23116, App
10	86.6	8.9	1312	US-10-216-817-21	Sequence 17, Appl
11	86.2	8.8	1425	US-10-437-963-1325	Sequence 23, Appl
12	82	8.4	9909	US-10-158-844-12	Sequence 1, Appl
13	81.2	8.3	968	US-09-864-866-46	Sequence 106, App
14	72.6	7.4	2256646	US-10-470-565-1	Sequence 106, App

15	71.4	7.3	930	US-10-156-761-4577	Sequence 4577, App
16	71.4	7.3	9025608	US-10-156-761-1	Sequence 1, Appl
17	68.2	7.0	991	US-10-398-221-1588	Sequence 1588, App
18	68.2	7.0	3155	US-10-398-221-3596	Sequence 3596, App
19	66.8	6.9	11058	US-10-156-761-3629	Sequence 3629, App
20	66.8	6.9	18876	US-10-329-079-42	Sequence 42, Appl
21	66.8	6.9	61944	US-10-329-079-34	Sequence 34, Appl
22	64.6	6.6	1230	US-10-156-761-5404	Sequence 5404, App
23	63.8	6.5	8244	US-10-402-843-1	Sequence 3, Appl
24	63.8	6.5	47968	US-10-402-843-1	Sequence 1, Appl
25	63.8	6.5	2428	US-10-402-843-1	Sequence 23116, App
26	62.4	6.4	2731748	US-10-297-465A-1	Sequence 1, Appl
27	62.4	6.4	32768	US-09-070-927A-17	Sequence 17, Appl
28	62.2	6.4	1237	US-10-458-201-23	Sequence 23, Appl
29	62.2	6.4	15120	US-10-458-201-23	Sequence 1, Appl
30	61.8	6.3	1949	US-10-437-963-50882	Sequence 50882, App
31	61.4	6.3	1230025	US-10-289-762-1	Sequence 1, Appl
32	61.2	6.3	975	US-09-938-901-1	Sequence 1, Appl
33	60.6	6.2	1041	US-10-437-963-8597	Sequence 8597, App
34	60.6	6.2	1484	US-10-437-963-29507	Sequence 29507, App
35	60.4	6.2	586	US-10-424-599-105239	Sequence 105239, App
36	60.4	6.2	2036	US-10-424-599-105239	Sequence 12666, App
37	59.8	6.1	1029	US-09-938-901-5	Sequence 5, Appl
38	59.8	6.1	5355	US-10-205-032-19	Sequence 19, Appl
39	59.8	6.1	60196	US-10-205-032-19	Sequence 1, Appl
40	59.4	6.1	272	US-09-864-408A-9041	Sequence 9041, App
41	58.8	6.0	1122	US-10-156-761-4667	Sequence 4667, App
42	58.8	6.0	1128	US-10-671-403-106	Sequence 106, App
43	58.8	6.0	1128	US-10-671-403-106	Sequence 106, App
44	58.8	6.0	1128	US-10-670-844-106	Sequence 106, App
45	58.8	6.0	1128	US-10-671-134-106	Sequence 106, App

#### ALIGNMENTS

RESULT 1  
US-09-938-901-1  
Sequence 1, Application US/09938901  
Publication No. US20030008291A1  
GENERAL INFORMATION:  
APPLICANT: Yokoyama Shigeyuki  
TITLE OR INVENTION: GENE ENCODING DNA REPAIR ENZYME  
FILE REFERENCE: PH-1261-US  
CURRENT APPLICATION NUMBER: US/09/938, 901  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: JP2001-47762  
PRIOR FILING DATE: 2001-02-23  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 975  
TYPE: DNA  
ORGANISM: Thermus thermophilus  
FEATURES:  
NAME/KEY: CDS  
LOCATION: (1)..(975)  
US-09-938-901-1

Query Match	99.9%	Score 974	DB 10	Length 975
Best Local Similarity	100.0%	Pred. No. 1.7e-209		
Matches 974	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	2	TGAGGCGCTGCGGAAAGCCCTCTCGCTGTA	CGGGAAGCCGCCCTCTCCCT	61
DB	2	TGAGGCGCTGCGGAAAGCCCTCTCGCTGTA	CGGGAAGCCGCCCTCTCCCT	61
QY	62	GCGCGGGGAGAGAGACCTTACCGCGCTCTCGAGTCTCCGAGTCTTCTTGAGAGAGACC		121
DB	62	GCGCGGGGAGAGAGACCTTACCGCGCTCTCGAGTCTCCGAGTCTTCTTGAGAGAGACC		121
QY	122	GCGGAGAGAGAGAGACCTTATTAACGCGCTTCTTGAGAGAGCTTCCACCTGAAG		181

```

Db      122 GGTGAGAGAGGCGCTTCCCTATTACCGCGCTTCTGTAGAGCGTTTCCACCTGAAAG 181
Qy      182 CCCTGCGCGCGCTTCCCTGGAAGAGTCTTAGGCTGAGGAGGCGGCTTACTCC 241
Db      182 CCCTGCGCGCGCTTCCCTGGAAGAGTCTTAGGCTGAGGAGGCGGCTTACTCC 241
Qy      242 GCGCGCGCGGAAACCTTCCACCGCTTGGCCCGGAAAGCTGTAGAGGCTTCCCGAGCTTGC 301
Db      242 GCGCGCGCGGAAACCTTCCACCGCTTGGCCCGGAAAGCTGTAGAGGCTTCCCGAGCTTGC 301
Qy      302 CCAGCTTGGGGGCGCTTCTGTGCTCGGCGCTTACACCGCGCGCGGCGGCTTCCACAG 361
Db      302 CCAGCTTGGGGGCGCTTCTGTGCTCGGCGCTTACACCGCGCGCGGCGGCTTCCACAG 361
Qy      362 CTTTGGGAGAGCGGCTGCGCGCGGTGAGACGAGAACGTCGAGAGGCTTCTCCGCGCTCT 421
Db      362 CTTTGGGAGAGCGGCTGCGCGCGGTGAGACGAGAACGTCGAGAGGCTTCTCCGCGCTCT 421
Qy      422 TCGCCCGGGAAGCCCGCAAGAGAGAGCTTTTGGCCCTTGGCCCGCGCGCTTCTCCCG 481
Db      422 TCGCCCGGGAAGCCCGCAAGAGAGAGCTTTTGGCCCTTGGCCCGCGCGCTTCTCCCG 481
Qy      482 AGGCGGTGAGACCGGCGGAGGTGAGAACAGGCGCTTATGAGAGCTGGGCGCACGCTCTGCC 541
Db      482 AGGCGGTGAGACCGGCGGAGGTGAGAACAGGCGCTTATGAGAGCTGGGCGCACGCTCTGCC 541
Qy      542 TCGCGAAACGCGCCCGCTTGGCGCGCGCTTCCCTTAGGAGCGCTTCTGCGCGGAGAGAG 601
Db      542 TCGCGAAACGCGCCCGCTTGGCGCGCGCTTCCCTTAGGAGCGCTTCTGCGCGGAGAGAG 601
Qy      602 CCGCGCGGCGCTACCGCGCGCGCTCAGAGAGCGCGGCGAGAGAGAGAGCGCTGCGCGC 661
Db      602 CCGCGCGGCGCTACCGCGCGCGCTCAGAGAGCGCGGCGAGAGAGAGAGCGCTGCGCGC 661
Qy      662 TCGCTCTCTCTGCGGAGAGAGGAGGTGACCTGTGAAAGCTTTGAGAGGCGCTTCCAGGCGC 721
Db      662 TCGCTCTCTCTGCGGAGAGAGGAGGTGACCTGTGAAAGCTTTGAGAGGCGCTTCCAGGCGC 721
Qy      722 TCTACGCGGCTCCCGCTTCTTCCCTGAGAGAGCTTCCGCGGCGGAGAGCGCGCTTGGCG 781
Db      722 TCTACGCGGCTCCCGCTTCTTCCCTGAGAGAGCTTCCGCGGCGGAGAGCGCGCTTGGCG 781
Qy      782 TGAAGTCTAAGCGCCCTTAGGCGAGGTGCGCACCGCGCTTACCCACCGAGAGCTTTCGCGTGG 841
Db      782 TGAAGTCTAAGCGCCCTTAGGCGAGGTGCGCACCGCGCTTACCCACCGAGAGCTTTCGCGTGG 841
Qy      842 AGGTGCGGAGGCGCTTGTGGAGAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
Db      842 AGGTGCGGAGGCGCTTGTGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
Qy      902 AGCTCATGAGAGAGTGTCTCGCAAGCGCTTCCCTCTCTCATGCGGCGGTGTAGTCC 961
Db      902 AGCTCATGAGAGAGTGTCTCGCAAGCGCTTCCCTCTCTCATGCGGCGGTGTAGTCC 961
Qy      962 CCGTCCCGGAGCGA 975
Db      962 CCGTCCCGGAGCGA 975

```

```

RESULT 2
US-10-156-761-4691
; Sequence 4691, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

```

```

; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4691
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(939)
US-10-156-761-4691

Query Match      11.5%; Score 112.4; DB 15; Length 939;
Best Local Similarity 52.1%; Pred. No. 3,1e-16;
Matches 369; Conservative 0; Mismatches 306; Indels 33; Gaps 4;

Qy      4 GAGGCGTGGGGAAGCCCTCTGCGCTGTACCGGGAAGAGCCCGCCCTTCCCTGCG 63
Db      73 GAGGCGTGGGGAAGCCCTCTGCGCTGTACCGGGAAGAGCCCGCCCTTCCCTGCG 132
Qy      64 GCGGG--GAGAGAGACCTTACCGCGTCTGCTGTCCAGAGTCTTCTGAGAGCTTCCACCTGAG 120
Db      133 CCGCCCGCGAGCGCGCGCCCTGAGAGGTGAGTGTGAGTGTATGTTGACACACACA 192
Qy      121 CCGGTGAGACAGAGCCCTTCCCTATTACCGCGCTTCTGAGAGCTTCTGAGAGCTTCCACCTGAG 180
Db      193 CCGGTGAGACAGAGCCCTTCCCTATTACCGCGCTTCTGAGAGCTTCTGAGAGCTTCCACCTGAG 252
Qy      181 GCGTGGCGCGGAGCTTCCCTGGAAGAGGTCTTGAAGGTGAGAGAGGAGGAGGAGCTTAC 240
Db      253 GACCTGCGCAAGAGAGGCGCGCGGAGAGCCCTCGCGCTGAGAGGAGGAGGAGGAGGAG 312
Qy      241 CCGGCGGAGGAACCTTACCGCGCTGAGC-----GGAAGCTGAGAGAG 285
Db      313 CCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 372
Qy      286 CTTCCCGGAGCTTACCGGAGCTTGGGAGGCTTCTGAGGCTTCTGAGGCTTCTGAGGCTTCTGAGG 345
Db      373 GTACCCAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 432
Qy      346 GCGGTGAGCTTCACTGCGCTTGTGAGAGCGGAGTGGCGCGGTGAGAGGAGAGAGGAGGAGG 405
Db      433 GCGGTGAGCTTCACTGCGCTTGTGAGAGCGGAGTGGCGCGGTGAGAGGAGAGAGGAGGAGG 492
Qy      406 GTCCCTTCCCGGCTTCTGCGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 459
Db      493 GTCTTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 552
Qy      460 -----CTGCGCAAGGAGCTTCTCCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 510
Db      553 CGAAGCTGCGCGCGCGGCTGCTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 612
Qy      511 GCGCTCATGAGAGCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 570
Db      613 GCGTCAATGAGAGCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 672
Qy      571 CCGCTAAGGAGCTTCTGCGCGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 630
Db      673 CCGATGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 732
Qy      631 CCGCGGAGGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 678
Db      733 CCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780

```

```

RESULT 3
US-10-156-761-1/c
; Sequence 1, Application US/10156761

```







```

; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 1325
; LENGTH: 5850
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_101202C.1
US-10-437-963-1325

```

```

Query Match          9.4%; Score 91.2; DB 17; Length 5850;
Best Local Similarity 54.4%; Pred. No. 1.2e-11;
Matches 211; Conservative 0; Mismatches 168; Indels 9; Gaps 1;

```

```

QY 65 GGGGGGAGAGACCTTACCGGCTCTGCTCCGAGGCTCTTTCGACGACAGCCCGG 124
DB 263 GCAGAGGAGAGAGGCGCTACGCGGTGTGGGTGTGGAGTGTGACGACATACGCGGG 322
QY 125 TGGAGCAGGCGCTCCCTATTACCGCGCTTTCTGAGCGCTTCCACCTGAGAGCC 184
DB 323 TCCCGGTGTGTGCTGACTACTCTCCGCTGTGATGCGCCCTGGCCACCGTGAAGCC 382
QY 185 TGGCGCGGCTTCTCTGAGAGAGTCTTGAAGGTCTGCAAGGGGCGGCTACTACCGC 244
DB 383 TCCCGGTGTGCTGACTACTCTCCGCTGTGATGCGCCCTGGCCACCGTGAAGCC 442
QY 245 GGGCGGAAACCTTCAACCGGCTGGCCGAGCGT-----GAGAGAGTTCCTCCCGA 295
DB 443 GGGCTCATTTCTTCTGAGAGGACAAAGCAATTTGAAAAAGCGAGTTCCTGCA 502
QY 296 GCTTCGCGAGCTTCGAGGAGCTTCTGCTCTGCGGCTTACACCGCGCGCGGTGCGCT 355
DB 503 CAGCATAGCACTCCGTAAGTTCGTGCAATGGGATTTACACGAGCGAGCCATTGCTT 562
QY 356 CCATCGCTTCGAGGAGCGGCTGCGCGGTGAGCGGAAAGTTCGAGAGGCTCTTCCC 415
DB 563 CCATAGCTTCAATGAGGTGTCTCTGTGTGATGAGAAATGTCGAGTTATCAGCAGC 622
QY 416 GCCTCTTCCCGGAGAAAGCCCAAGA 443
DB 623 TTACGCTATTCTGATTAACCAAGA 650

```

```

RESULT 8
US-10-425-114-32132
; Sequence 32132, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovacic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53131)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 2003-04-28
; SEQ ID NO 32132
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Zea mays

```

```

; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73265D07_FLI
US-10-425-114-32132

```

```

Query Match          9.1%; Score 89.2; DB 13; Length 1660;
Best Local Similarity 50.7%; Pred. No. 4.4e-11;
Matches 282; Conservative 0; Mismatches 253; Indels 21; Gaps 2;

```

```

QY 60 CTGGCGGGGAGAGAACCTTACCGGCTCTGCTCTCCGAGTCTCTGAGAGAGC 119
DB 278 CGGACGAGAGAGAGGCGTACCGGCTGTGGGTGTCCGAGTATCTGACAGAGC 337
QY 120 CGGGGTGAGAGAGGCGCTCCCTATTACCGCGCTTCTGAGCGCTTCCACCTGAA 179
DB 338 GCGGCTGCGCGGTGTGTGCTTACTACAGCGGTGATGCGCGGTGCGCCACCTAGC 397
QY 180 GGCCTGAGCGCGGCTTCCCTGAGAGAGTCTTGAAGTCTGAGAGGCGGCTACTA 239
DB 398 AGCCTGCGCGCGCTACCGAGAGAGGTGAACGAGATGTGGGCGGCTTGGCTACTA 457
QY 240 CCGGCGGGGAGAACACTCC-----ACGCGTGGCCGAGAGGTGAGAGCTTCC 290
DB 458 CCGTGGGCTCGATTTCTGCTGAGAGAGCAAGCATATCGAAAGGGGTTGTTCC 517
QY 291 CCGAGCTTCCGAGCTTCGAGGCTTCGAGGCTCTGAGTCTGAGGCTTACACCGCGCGGT 350
DB 518 TTGACAGCGGTTCGCTTCTGAGGTGTGAGTGTGAGATTTACAGCTGAGCAT 577
QY 351 GGCCTCATCGCTTCGAGGAGCGGCTGAGCGGCTGAGAGCGTTCGAGGCTCT 410
DB 578 CGCTTCAATTCCTTCAAGAGAGTGTCCAGTTGTGATGAGAAATGTGATACAGTCA 637
QY 411 CTCCCGCTTCTTCCCGGAGAAAGCCCAAGAGAA-----GAGCTTTCCG 458
DB 638 CAGCAGGCTTTACACCATTTCTGACAAACCAAGAGATCTCAACAGTGAAGATTCG 697
QY 459 CTTGCGCGAGGCTCTCTCCCGAGGCGGTGAGACCGGAGGTGTGAAACGAGCCCTCAT 518
DB 698 GGACTGTGATGATCAATATGTTTACCTTTGAGACAGAGAGATTTCAACCAACGATGAT 757
QY 519 GAGCTCGGGGACAGGTCTGCTGCGGAAACGCGCCGCTTGGGGGCTGCCCCCTAG 578
DB 758 GAGTTAGAGAGCAATTAATGTAAGCAAGAAAGCTGTTGCTCAATGCCCCAGTCTC 817
QY 579 GGCCTTCTCGGCGG 594
DB 818 TAGCCACTGCGCAAGCG 833

```

```

RESULT 9
US-10-216-817-31
; Sequence 31, Application US/10216817
; Publication No. US20030129619A1
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
; FILE REFERENCE: 03495.0233-00000
; CURRENT APPLICATION NUMBER: US/10/216,817
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 60/311,824
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,523
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
; LENGTH: 912
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-216-817-31

```

Query Match 8.9%; Score 86.6; DB 15; Length 912;  
 Best Local Similarity 50.2%; Pred. No. 1.9e-10;  
 Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

```

QY 22 CTCCTCCGCTGTACCGGAAAAAGCCGCCCTCCCTGCGGGG---GAGAAGAC 78
DB 61 CTTCTCGGTTGGTATACAGATCCGATCCGAGATCTGCCCTGGAGACCCGGTGTACG 120
QY 79 CCTTACCGCGTCTCTGAGAGTCTTTCGAGAGAGACCCGGGTGAGACAGCCCTC 138
DB 121 CCGTGGCAGATCTGTGTACGAGATTCATGCTGAGAGACCGCCGCGGATGCTG 180
QY 139 CCTTATTAACCGCCGCTTTCTGAGCGCTTTCACCCCTGAAGGCCCTGGCCGCTTCC 198
DB 181 GCGATCTGGCCGAGCTGGAGTGGGGGCGCCACGCCGTGGCCACCGCCAGGCCAGC 240
QY 199 CTGAAGAGGTCTTAAAGGTCTTGACAGAGGGGCGGGCTTACTACCGCGGGGGAACACTC 258
DB 241 ACCGCCGATGTGTATACGCGCTTGAGGCAAGCTGGGCTATCCAGCGCAGCCAGCCTTA 300
QY 259 CACCGCCTGGCCCGAAGCGTG-----GAGAGCTTCCCGAGCTTGCC 303
DB 301 CACGAGTGGCCACCGTATGCGCCCGGACCAATAGCTGTGCTCCGACGATATGAG 360
QY 304 GAGCTTGGGGGCTTCTGTGCTTGGGCTTACACCGCGGCGGAGTGGCTTCATGCC 363
DB 361 ATCTGTGTCACCCCTGCGGGGCTGGAGAGCTACCCGCGCGGTGGTGTTCGCT 420
QY 364 TTCGGGAGCGGAGTGGCGGAGTGAAGAGGAGTCTCCGAGGGTCTTCCGCTT- 422
DB 421 TACCGCAAGCGGGTGGCGGTGTGACACCAATGTGCGGCGGTGGCGCGCGCTT 480
QY 423 -----CGCCGAGAAAGCCCAAGAGAGAGCTTTTCGCTCGCCCAAGGC 471
DB 481 CACGCGCCCGCAGCGCGGTGGCGCATCGGTGCGGCGACACGCGCAAGCTTGGCG 540
QY 472 CTCCTCCCGAGGCGGTGAGACCGGGGGGTGGAACCAAGGCCCTCATGAGCTCGAGCC 531
DB 541 CTGTGTCCGACCGCGAGACGCGCGCTGAATTTTCGCTCGCTGATGAGTGGGTGGC 600
QY 532 ACGGTCTGCTGCGGAACCGGCCCTGTGCGGGGCTGCGCCCTGAAGGGGCTTCTGCGCG 591
DB 601 ACGGTGTGACACCGCCGCAACCCCGGTGCGGTATGCGCGTGAAGTGTGCGATGG 660
QY 592 GGAAGAGAGCGCCCGGCGCTAACCCGCGCCCAAGAGCGCGGCGGAAGAGAGAGCGC 651
DB 661 CGGCAATGCCGTTATCCGCGTGGAGCGGTCC---GCCGCGCGGGGGGAGGCTTACAC 717
QY 652 CTCGTGCGCTCGTCTCTCTCGGCGG 678
DB 718 GGAACGAGCGCCAGTCCGCGAGCG 744

```

RESULT 10  
 US-10-216-817-21  
 ; Sequence 21, Application US/10216817  
 ; Publication No. US20030129619A1  
 ; GENERAL INFORMATION:

APPLICANT: BRIGITTE  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
 TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
 FILE REFERENCE: 03495-0233-00000  
 CURRENT APPLICATION NUMBER: US/10/216,817

PRIOR FILING DATE: 2002-11-13  
 PRIOR APPLICATION NUMBER: 60/311,824  
 PRIOR FILING DATE: 2001-08-14  
 PRIOR APPLICATION NUMBER: 60/113,523  
 NUMBER OF SEQ ID NOS: 32

SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 21  
 LENGTH: 1312

TYPE: DNA  
 ORGANISM: Mycobacterium tuberculosis  
 US-10-216-817-21

Query Match 8.9%; Score 86.6; DB 15; Length 1312;  
 Best Local Similarity 50.2%; Pred. No. 1.8e-10;  
 Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

```

QY 22 CTCCTCCGCTGTACCGGAAAAAGCCGCCCTCCCTGCGGGG---GAGAAGAC 78
DB 261 CTTCTCGGTTGGTATACGAGATCCGAGATCTGCCCTGGAGACCCGGTGTACG 320
QY 79 CCTTACCGCGTCTCTGAGAGTCTTTCGAGAGAGACCCGGGTGAGACAGCCCTC 138
DB 321 CCGTGGCAGATCTGTGTACGAGATTCATGCTGAGAGACCGCCGCGGATGCTG 380
QY 139 CCTTATTAACCGCCGCTTTCTGAGCGCTTTCACCCCTGAAGGCCCTGGCCGCGCTTCC 198
DB 381 GCGATCTGGCCGAGCTGGAGTGGGGGCGGCCACGCCGTGGCCACCGCCAGGCCAGC 440
QY 199 CTGAAGAGGTCTTAAAGGTCTTGACAGAGGGGCGGGCTTACTACCGCGGGGGAACACTC 258
DB 441 ACCGCCGATGTGTATACGCGCTTGAGGCAAGCTGGGCTATCCAGCGCAGCCAGCGCTTA 500
QY 259 CACCGCCTGGCCCGAAGCGTG-----GAGAGCTTCCCGAGCTTGCC 303
DB 501 CACGAGTGGCCACCGTATGCGCCCGGACCAATAGCTGTGCTCCGACGATATGAG 560
QY 304 GAGCTTGGGGGCTTCTGTGCTTGGGCTTACACCGCGGCGGAGTGGCTTCATGCC 363
DB 561 ATCTGTGTCACCCCTGCGGGGCTGGAGAGCTACACCGCGCGGTGGTGTTCGCT 620
QY 364 TTCGGGAGCGGAGTGGCGGAGTGAAGAGGAGTCTCCGAGGGTCTTCCGCTT- 422
DB 621 TACCGCAAGCGGGTGGCGGTGTGACACCAATGTGCGGCGGTGGCGCGCGCTT 680
QY 423 -----CGCCGAGAAAGCCCAAGAGAGAGCTTTTCGCTCGCCCAAGGC 471
DB 681 CACGCGCGCGCGAAGCGCGGTGGCGCATCGGTGCGGCGACACGCGCAAGCTTGGCG 740
QY 472 CTCCTCCCGAGGCGGTGAGACCGGGGGGTGGAACCAAGGCCCTCATGAGCTCGAGCC 531
DB 741 CTGTGTCCGACCGCGAGACGCGCGCTGAATTTTCGCTCGCTGATGAGTGGGTGGC 800
QY 532 ACGGTCTGCTGCGGAACCGGCCCTGTGCGGGGCTGCGCCCTGAAGGGGCTTCTGCGCG 591
DB 801 ACGGTGTGACACCGCCGCAACCCCGGTGCGGTATGCGCGTGAAGTGTGCGATGG 860
QY 592 GGAAGAGAGCGCCCGGCGCTAACCCGCGCCCAAGAGCGCGGCGGAAGAGAGAGCGC 651
DB 861 CGGCAATGCCGTTATCCGCGTGGAGCGGTCC---GCCGCGCGGGGGGAGGCTTACAC 917
QY 652 CTCGTGCGCTCGTCTCTCTCGGCGG 678
DB 918 GGAACGAGCGCCAGTCCGCGAGCG 944

```

RESULT 11  
 US-10-437-963-35788  
 ; Sequence 35788, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:

APPLICANT: LA ROSA, Thomas J.  
 APPLICANT: KOVALIC, David K.

APPLICANT: ZHOU, Yihua  
 APPLICANT: CAO, Yongwei

APPLICANT: BUKHAROV, Andrey A.  
 APPLICANT: BARBAZUK, Brad

APPLICANT: LI, Ping

TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 35788  
 ; LENGTH: 1425  
 ; TYPE: DNA  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_39677C.1  
 US-10-437-963-35788

Query Match 8.8%; Score 86.2; DB 17; Length 1425;  
 Best Local Similarity 55.7%; Pred. No. 2.1e-10;  
 Matches 191; Conservative 0; Mismatches 143; Indels 9; Gaps 1;

65 GGGGGGAGAGAGACCTTACCGGCTCTCCAGGTCTCTTCCAGACAGACCGG 124  
 266 GCGAG 325  
 125 TGGAG 184  
 326 TACCGGAG 385  
 185 TGGCGGAG 244  
 386 TCGCGGAG 445  
 245 GGGGGGAG 295  
 446 GGGGTCAGATTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 505  
 296 GCTTCCGAG 355  
 506 CAGCATCAACATCTCGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 565  
 356 CCATCGCTTCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 398  
 566 CCAAGCCTTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 608

RESULT 12  
 US-10-158-844-12/c  
 ; Sequence 12, Application US/10158844  
 ; Publication No. US20040029118A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kunesh et al.  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 391  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: CD-R  
 ; COMPUTER: Dell Latitude Pentium 3  
 ; OPERATING SYSTEM: Windows 98  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/158,844  
 ; FILING DATE: 03-Jun-2002  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/961,527  
 ; FILING DATE: 1997-10-30  
 ; APPLICATION NUMBER: US 60/029,960  
 ; FILING DATE: 1996-10-31  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Hyman, Mark J.  
 ; REGISTRATION NUMBER: 46,789  
 ; REFERENCE/DOCKET NUMBER: PB340P1D1

; INFORMATION FOR SEQ ID NO: 12:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 9909 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
 US-10-158-844-12

Query Match 8.4%; Score 82; DB 13; Length 9909;  
 Best Local Similarity 52.2%; Pred. No. 1.3e-09;  
 Matches 218; Conservative 0; Mismatches 185; Indels 15; Gaps 1;

22 CTCCTGCGCTGATACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 81  
 5808 CTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5749  
 82 TACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 141  
 5748 TATCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5689  
 142 TATTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 201  
 5688 TACTACGAGAGATTTTGGAGCTGTTCCAACTGTGAGAAAGTCTGCAACTG 5629  
 202 GAAGAGTCTTGGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 246  
 5628 GAGAGTTTACTGAAACCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5569  
 247 GCGGAGACCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306  
 5568 GCTGCAAG 5509  
 307 CTTGCGGAG 366  
 5508 ATTTCAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5449  
 367 GGGAG 424  
 5448 AACTTGCTGAG 5391

RESULT 13  
 US-09-864-866-46  
 ; Sequence 46, Application US/09864866  
 ; Patent No. US20020127656A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lloyd, R. Stephen  
 ; APPLICANT: McCullough, Amanda K.  
 ; TITLE OF INVENTION: DNA REPAIR POLYPEPTIDES AND METHODS OF USE  
 ; FILE REFERENCE: 265,00170101  
 ; CURRENT APPLICATION NUMBER: US/09/864,866  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/206,279  
 ; PRIOR FILING DATE: 2000-05-23  
 ; NUMBER OF SEQ ID NOS: 49  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 46  
 ; LENGTH: 968  
 ; TYPE: DNA  
 ; ORGANISM: Micrococcus luteus  
 US-09-864-866-46

Query Match 8.3%; Score 81.2; DB 9; Length 968;  
 Best Local Similarity 47.2%; Pred. No. 3e-09;  
 Matches 300; Conservative 0; Mismatches 318; Indels 18; Gaps 1;

44 AGCG 103  
 200 ACCGTACG 259  
 104 TCCTTTCGAGCAACCGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 163



QY 321 TGGTCTCGGGCCCTTAACCGCGGGCGGTGSCCTTCATGCTTGGGGAGCGGCTGGC 380  
 Db 507 CGGCGTAGGACGCAGACGGCTTCGTGGTGTCTGGCAACGCTTCGGGCGGCCGACAT 566  
 QY 381 GGGGTGAGCGGGAAGTCCGGAGGGTCTTCGCCCTCTTGGCCCGGAAAGCCCA 440  
 Db 567 CACGTGAGCACCATTTCCAGCGGCTGTCGCCGCTGGCAGTGGACCGAGAGAGA 626  
 QY 441 GGAGAGAGAGCTTTGCGCCCTCGCCAGGGCTCTCCCGAGAGGCTGAGCCCGGGGT 500  
 Db 627 TCCGACAGATCGAGCGGCGCTTCGGCGGCTCTTCCGAGAGGAGTGAACGATGCT 686  
 QY 501 GTGAAACAGGCGCTCATGAGCTCGGGGCGACGCTCTGCGCGAAGCGCCCGTTG 560  
 Db 687 CTCGACCATG--TGATCTTCCACGGCGCGCATCTGTCAAGCCCGCAAGCGCGCTG 743  
 QY 561 CGGGGCTGCGCCCTTAGGGGCTTTCGCCGGGGAGAGGCCCGCGGCGCTAACCCGC 620  
 Db 744 CGGGGCTGCGCCCATGCGCCCGCTCTGCGCGGCTAGGGGAGGGCGAGACCGACCCGA 803  
 QY 621 GCCCAGAGCGCGCGGAGAGAGGAGCGCTGTCGCCCTGCTCC 667  
 Db 804 GAAGGCGAGAGAGCTCTGAGTACGAGAGGGCGCTTCCCGGCC 850

Search completed: October 8, 2004, 06:39:54  
 Job time : 591 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 4, 2004, 03:57:10 ; Search time 2393 Seconds

(without alignments)  
12166.998 Million cell updates/sec

Title: US-09-938-901-1

Perfect score: 975  
Sequence: 1 gtgagagccgcggcgaaagc.....tagtcccccccgagcga 975Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rnd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	10.7	1014	12	BM561672
2	101.4	10.4	1056	10	BM561672 AGENCOURT
3	100.2	10.3	496	12	BF974701 602245343
4	99.4	10.2	969	13	BG383035 300920 MA
					BX367285 BX367285

5	95.2	9.8	1009	12	BM925099
6	91.6	9.4	964	13	BQ686151
7	90.6	9.3	662	12	B1818417
8	90.2	9.2	577	12	BG177663
9	90	9.2	771	12	B1818560
10	90	9.2	1014	13	BX401779
11	90	9.2	1933	11	AK077546
12	88.6	9.1	1031	12	BM921263
13	87.4	9.0	900	12	BG761186
14	85.8	8.8	982	13	BX415111
15	84.4	8.7	695	28	B2892784
16	84.2	8.6	563	29	CG898797
17	83.2	8.5	571	29	CG898763
18	82.6	8.5	999	13	BQ685480
19	82.4	8.5	1201	9	AL546337
20	82.2	8.4	1072	12	BM925133
21	81.2	8.3	935	29	BQ685409
22	81	8.3	935	29	CNS006XK
23	80	8.2	1179	12	BM549153
24	79.4	8.1	556	29	CG898299
25	79.2	8.1	982	13	BX415111
26	79	8.1	641	12	BQ224720
27	78.4	8.0	360	12	B1343166
28	78.4	8.0	530	10	BF075702
29	78.4	8.0	532	10	BE030801
30	78.4	8.0	551	10	AW656251
31	78.2	8.0	1203	29	CNS015Y4
32	77.8	8.0	925	29	CNS0091P
33	77.6	8.0	932	29	B1656309
34	77.6	8.0	932	29	CNS00720
35	77.2	7.9	786	14	CK021799
36	77.2	7.9	1203	29	CNS015Y4
37	76.6	7.9	828	13	BX425797
38	76	7.8	1083	14	CK206171
39	75.4	7.7	769	12	BG918941
40	75.4	7.7	865	13	BQ890365
41	74.6	7.6	563	12	BM830888
42	74.4	7.6	580	12	BM759737
43	74.4	7.6	594	12	BM745229
44	74.4	7.6	594	12	BM745229
45	74.4	7.6	981	12	BM926983

## ALIGNMENTS

RESULT 1  
LOCUS BM561672 1014 bp mRNA linear EST 20-PEB-2002  
DEFINITION AGENCOURT\_6593278 NIH\_MGC\_106 Homo sapiens CDNA clone IMAGE:5464344  
5', mRNA sequence.  
ACCESSION BM561672  
VERSION BM561672.1 GI:18807115  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1014)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-romail.nih.gov  
Tissue Procurement: Dr. Daniel McVicar, DMS/NCI  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L10M2013 row: h column: 09  
High quality sequence stop: 617.

FEATURES  
source

Location/Qualifiers  
1.1014  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:548434"  
/tissue\_type="natural killer cells, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_106"  
/note="Organ: Blood; Vector: pOTB7, Site\_1: XhoI, Site\_2: EcoRI, cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 10.7%; Score 104; DB 12; Length 1014;  
Best Local Similarity 54.0%; Pred. No. 5.6e-06;  
Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;

69 GGAGAAGACCTTACCGGCTCTGCTCCGAGTCTTTCGACAGACCCGGGTGA 128  
15 GGACAGCGGGCATATGCTGTGGGTCTCAGAGTCATGTCAGACAGCCAGATTGC 74  
129 GCAGGCGCTCCCTATTACCGCGCTTTTCGAGCGCTTTCCACCTGAAGCCCTGAC 188  
75 CACTGTGATCACTACTATACCGAGTGAATGACAGAGTGGCCCTACCTCAGACCTGGC 134  
189 CGGCGCTTCCCTGGAAGAGGCTCTTAGGCTGTCGACGGGGGGGCTACTACCGGGGGGC 248  
135 CAGTGTCTTCCCTGGAAGAGGCTGAATCACTTGGGCTGGCTGGCTACTATTCTGTGG 194  
249 GGAACACCTTCACCGCTGGCCCGAAGCGTGAAGAGCTTCCCGAGCTTC----- 300  
195 CGGCGGCTGACAGAGAGGAGCTCGAAGAGTGAAGAGAGGAGGCGCATGTCACAG 254  
301 -----GCCAGCTTGGGGGCTTCTGTCTCGGAGCTTACACCGCGCGGT 350  
255 TACAGCAGAGACCTTCGACAGCTCTGCTGGCGGCGCTACACAGCTGGGGCCAT 314  
351 GGCCTTCATGCTTGGGGAGCGGAGTGGCGGCGGTGAAGAGTCCGAGAGGCTCT 410  
315 TGCCTTATGCTTGGGCGAGCAACCGGTGTGATGAGTGAAGTGAACAGGATGCT 374  
411 CTCGCGCTCTTTCGCCC--GGAAAGCCCAAGAGAGAGAGCTTTTCCTCGCCAG 468  
375 GTGCGGTGTCGAGGCGATTGTGTGATCCAGACAGACCTTGTTCACAGAGCTTG 434  
469 GGCCTTCCTCCGAGAGGCGTGAACCG-----GGGTGTGAACAGAGCCCTCAT 518  
435 GGGTCTAGCCCAAGAGCTGTGAGCCAGACCCGGCCAGATTTCAACCAAGACCAT 494  
519 GGAGCTCGGGGCAAGGCTGTGCTGCGCAAGAGCGCCCTTACGGGGGCTGCCCCCTAG 578  
495 GGAAGCTAGGGCCACAGTGTGATCCCAAGCGCCCACTGTGACAGCAATGCTGTGGA 554  
579 GGCCTTTCGCGGAG 592  
555 GAGCCTGTGCGGAG 568

RESULT 2  
BF974701 1056 bp mRNA linear EST 22-JAN-2001  
LOCUS 602245343F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4336498 5',  
DEFINITION mRNA sequence.  
ACCESSION BF974701  
VERSION BF974701.1 GI:12341916  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 1056)  
NIH-MGC <http://mhc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL  
COMMENT  
Contact: Robert Stransberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: Incyte Genomics, Inc.  
National Gene Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1CM1209 row: e column: 11  
High quality sequence stop: 678.

FEATURES

source

Location/Qualifiers  
1.1056  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4336498"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7, Site\_1: XhoI, Site\_2: EcoRI, cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 10.4%; Score 101.4; DB 10; Length 1056;  
Best Local Similarity 53.7%; Pred. No. 1.4e-05;  
Matches 299; Conservative 0; Mismatches 226; Indels 32; Gaps 3;

69 GGAAGAAGACCTTACCGGCTCTGCTCCGAGTCTTTCGACAGACCCGGGTGA 128  
228 GGACAGCGGGCATATGCTGTGGGTCTCAGAGTATGTCGACAGACCCAGATTGC 287  
129 GCAGGCGCTCCCTATTACCGCGCTTTTCGAGCGCTTTCCACCTGAAGCCCTGAC 188  
288 CACTGTGATCACTACTATACCGAGTGAATGACAGAGTGGCTACCTGACAGAGCTGGC 347  
189 CGCGGCTTCCCTGGAAGAGTCTTAGGCTGTGACAGGGCGGGCTACTACCGGGGCG 248  
348 CAGTGTCTTCCCTGGAAGAGTGAATCACTTGGGCTGGCGCTGAGCTACTATCTGTGG 407  
249 GGAACACCTTCACCGCGCTGCGCCGAAAGCTGAAGAGCTTCCCGAGCTTC----- 300  
408 CCGGCGGCTGAGAGAGGAGCTCGAAGGTGTGAAGAGCTAGGGGCCACATGCCAG 467  
301 -----GCCAGCTTGGGGGCTTCTGCTGCTGGGCGCTTACACCGCGGCGGG 349  
468 TACAGCAGAGACCTTTCACAGAGCTCTGCTGGCGGTGGGCGCTTACAGCTGGGCCA 527  
350 TGGCTTCATGCTTGGGGAGCGGAGTGGCGCGGTGAACGGGAAGTCCGAGAGGCTCC 409  
528 TTGCTCTATGCTTTTGGCAGGCAACGGGTGTGATGAGTGAAGTGAACAGGGTGC 587  
410 TCTCCGCTCTTTCGCCC--GGAAAGCCCAAGAGAGAGCTTTTCGCGCCCA 467  
588 TGTGCGGTGTCCAGGCGATTGTGTGATCCAGACAGACCTTGTTCACAGAGCTCTG 647  
468 GGCCTTCTTCCCGAGGCGTGAACCC-----GGGGGTGTGAACAGAGCCCTC 516  
648 GGGTCTAGCCAGCAGCTGTGTGAGCCAGCGCCGCAAGAGATTTCAACCAAGCAGAC 707



QY 517 ATGAGCTGGGGGACGCTGCTGCTCCGAAGGCCCGCTTGGGGGCTGCCCCCTA 576  
 DB 708 ATGAGCTTACGAGGCGGACGAGGCTGACCCCAAGAGCCCACTGGGCGAGCCAGCTGCTGG 767  
 QY 577 GGGGCTTGTGCGGGG 593  
 DB 768 TGGAGAGCTGTGCGGG 784

RESULT 3  
 BG383035 496 bp mRNA linear EST 12-MAR-2001  
 LOCUS 3009920 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.  
 ACCESSION BG383035  
 VERSION BG383035.1 GI:13307507  
 KEYWORDS EST.  
 SOURCE Sus scrofa (pig)  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 1 (bases 1 to 496)  
 Fahrenkrug,S.C., Smith,T.P.L., Freking,B.A., Cho,J., White,J., Vallet,J., Wise,T., Rohrer,G.A., Pertee,G., Sultana,R., Quackenbush,J. and Keele,J.W.  
 Quackenebush,J. and Keele,J.W.  
 Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly  
 Mamm. Genome 13 (8), 475-478 (2002)  
 22213789  
 12226715

TITLE Contact: Smith TPL  
 JOURNAL USDA, ARS, US Meat Animal Research Center  
 MEDLINE PO Box 166, Clay Center, NE 68933-0166, USA  
 PUBMED Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@meatlib.mars.usda.gov  
 Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -mismatch 12 options.  
 PCR primers  
 FORWARD: AGGAACAGCATGACCAT  
 BACKWARD: GTTTCACGATCAGCAGC  
 Plate: 6 row: M column: 13  
 Seq primer: ATTTAGGTGACATTAAG.  
 Location/Qualifiers  
 1..496  
 /organism="Sus scrofa"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9823"  
 /rname\_type="pooled"  
 /lab\_host="DH10B"  
 /clone\_lib="MARC 1P1G"  
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."

ORIGIN  
 Query Match 10.3%; Score 100.2; DB 12; Length 496;  
 Best Local Similarity 57.1%; Pred. No. 1.8e-05;  
 Matches 214; Conservative 0; Mismatches 143; Indels 18; Gaps 1;

QY 250 GAACACTCCAGCGCTGCGCCCGAAGCGTGAAGAGCTTCCCGAGCTTC----- 300  
 DB 231 CGCTGCTTACAGACAGAGCGCCCGGAAGCTGTGAGAGAGCTAGAGGCCACATCCACGA 290  
 QY 301 -----GCCGAGCTTGGGGGCTTCTGTGCTCGGGCTTACACCGGCGCGGCTG 351  
 DB 291 ACAGACAGAGACCCCTGACAGCGGCTTCCTGCGCGTGGCGGCTGACAGCGCGGCGCAT 350  
 QY 352 GCTTCATCGCTTGGGAGACCGGCTGCGCGCTGAGACGGAACTCCGAGGGCTTC 411  
 DB 351 GCTTCATTCCTTTTGGCAGGAGCTGTGTGTGAGACGGAATGATTTGGGTGCTG 410  
 QY 412 TCCGCGCTTTCGCC 426  
 DB 411 TCCGCGCTTCGAGGC 425

RESULT 4  
 BX367285 969 bp mRNA linear EST 08-MAY-2003  
 LOCUS BX367285 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 DEFINITION Homo sapiens cDNA clone CS0DL004YB24 5'-PRIME, mRNA sequence.  
 ACCESSION BX367285  
 VERSION BX367285.1 GI:30449616  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 969)  
 Li,W.B., Gruber,C., Jesse,J. and Polayres,D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seq@genoscope.cns.fr; Web: www.genoscope.cns.fr  
 Library was constructed by life Technologies, a division of invitrogen. This sequence belongs to sequence cluster 922.r for more information about this cluster. see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DL004DA12QPI&cluster=922.r. Contact :  
 Peng Liang Email: filiang@lifetech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DL004DA12QPI.  
 Location/Qualifiers  
 1..969  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DL004YB24"  
 /cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
 Query Match 10.2%; Score 99.4; DB 13; Length 969;  
 Best Local Similarity 53.1%; Pred. No. 2.7e-05;  
 Matches 296; Conservative 2; Mismatches 228; Indels 31; Gaps 3;

QY 189 CGCGGCTTCCCTGGAAAGAGCTCTTAGAGGCTCTGGACAGGGGGGGGGCTACTACCGCGGCG 248  
 Db 521 CAGTGCCTTCCCTGGAGAGAGTGAATCAACTTGGGCTGGCGCTGAGCTACTATTCTCGTGG 580  
 QY 249 GGAAACACTTCACCGCGCTGGCCCCGAAAGCGTGAAGAGACTTTCGCCGAGCTTC----- 300  
 Db 581 CCGGCGGCTGACAGAGAGGAGACTCGAAAGAGTGAAGAGACTAAGGGGGCCACATGCCAG 640  
 QY 301 -----GCCAGCTTCGGGAGGCTTCTGTGCTCTCGGCGCTTACACCGCGGCGGAGT 350  
 Db 641 TACAGCAGAGACCTTGACAGACTCTGCTGCTGGCGTGGGCGCTACACAGCTGGGGGCAT 700  
 QY 351 GGCTTCATGCGCTTCCGGGAGCGGGTGGCGGCGTGAACGGGAAACGTCCGAGAGGCTT 410  
 Db 701 TGCCTCATAGCGCTTTGGCCAGGCAACCGGTGTGTGATGGCAACGTAGCAGCGGAGCT 760  
 QY 411 CTCCGCGCTTTCGCCC--GGGAAAGCCCCAAGAGAAAGAGCTTTTGCCCTTCGCCAG 468  
 Db 761 GTGCGCGGTCCGAGCAATTGTGCTGATCCAGCAGACCTTGTTTCCAGCAGGCTCTG 820  
 QY 469 GGCCTCTCTCCGAGGGGCGTGAACC-----GGGGGTGGAACCAAGCCCTCA 517  
 Db 821 GGGCTAGCCCAAGCACTGTGTGAACCAAGCCCGCCAGAGGAGATTTCAACCAAGACCA 880  
 QY 518 TGAAGCTCGGAGGCAACAGTCTGCTGCGGAAACGGCCCCGCTTGGCGGAGCTTGCCCCCTAG 577  
 Db 881 TGAAGCTAGGGGCAACAGTGTGTAACCCCAAGCGCCCACTGTGACGCACTKCTCTTGG 940  
 QY 578 GGGCCTTCGCGCGGGG 594  
 Db 941 AGAGCTGTTCCGGGAG 957

FEATURES	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
LOCUS	BM925099				
DEFINITION	AGENCYCOURT_6627519 NIH_MGC_122 Homo sapiens				
ACCESSION	BM925099				
VERSION	BM925099.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
	1 (baae 1 to 1009)				
	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
	National Institutes of Health, Mammalian Gene Collection (MGC)				
	Unpublished (1999)				
	Contact: Robert Struhsberg, Ph.D.				
	Email: <a href="mailto:cgagbbs@mail.nih.gov">cgagbbs@mail.nih.gov</a>				
	Tissue Procurement: Life Technologies, Inc.				
	cDNA Library Preparation: Life Technologies, Inc.				
	DNA Sequencing by: Agencourt Bioscience Corporation				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LNL at:				
	<a href="http://image.lnl.gov">http://image.lnl.gov</a>				
	Plate: LRAM2812				
	High quality sequence stop: 655.				
	Location/Qualifiers				

```

FEATURES
Source
Location/Qualifiers
1. 1009
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5762392"
/lab_host="DH10B"
/clone_11b="NH_WGC_122"
/note="Organ: pooled lung and spleen; Vector: pCMV-Sport6
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 34 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed

```

upon cloning. Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note this is a NIH\_MGC Library."

Query Match	9.8%	Score 95.2;	DB 12;	Length 1009;
Best Local Similarity	53.3%;	Pred. NO. 0.00012;		
Matches 289; Conservative	0;	Mismatches 223;	Indels 30;	Gaps 3;

[illegible]

RESULT 6	BO666151	LOCUS	BO666151	DEFINITION	BO666151	964 bp	mRNA	linear	EST 15-JUL-2002
ACCESSION	AG000001	VERSION	8344868	KEYWORDS	NT_H_MGC_110 Homo sapiens	CDNA clone IMAGE:6247891			
SOURCE	EST.	ORGANISM	Homo sapiens (human)						
REFERENCE	1 (baee 1 to 964)	TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)						
AUTHORS	Unpublished (1999)	JOURNAL	Contact: Robert Strauberg, Ph. D.						
COMMENT	Email: cga@bs-romail.nih.gov		Tissue Procurement: ATCC						
	CDNA Library Preparation: Rubin Laboratory		CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)						



RESULT 8  
BG177663 577 bp mRNA linear EST 06-FEB-2001  
LOCUS 602314663F1 NIH\_MGC\_85 Homo sapiens CDNA clone IMAGE:4420430 5',  
DEFINITION mRNA sequence.  
ACCESSION BG177663  
VERSION BG177663.1 GI:12684366  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
1 (bases 1 to 577)  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgabbs-remail.nih.gov  
Tissue Procurement: Louis Straudt, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.jnl.gov  
Plate: LLM0157 row: n column: 15  
High quality sequence stop: 577.

FEATURES  
source  
1. 577  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4420430"  
/issue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC 85"  
/note="Organ: lymph; Vector: PCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; cloned unidirectionally; oligo-dT primed.  
Average insert size 1.867 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 9.2%; Score 90; DB 12; Length 577;  
Best Local Similarity 55.3%; Pred. No. 0.00062;  
Matches 208; Conservative 0; Mismatches 150; Indels 18; Gaps 1;  
QY 69 GGAGAGACCTTACCGCGCTCTGCTCCGAGGCTCTTTCGACAGACCCGGGTGGA 128  
DB 186 GGACAGCGGGCATATGCTGTGTGAGGTCTGACAGGTCATGACAGACCCAGGTTGC 245  
QY 129 GCAGGCGCTCCCTATTACCGCGCTTTCTGAGAGCGCTTCCACCCCTAAAGCCCTGAC 188  
DB 246 CACTGTGATCACTACTATACCGATGATGAGAGAGAGGCTTACACAGGACCTTGCC 305  
QY 189 CGCGGCTTCTGAGAGAGGCTTTCAGGCTTCGACAGGCGGCGCTACTACCGCGGAC 248  
DB 306 CAGTGTCTCCCTGAGAGAGGTGAATCACTCTGGGCTGGCCCTGAGTACTATTCCTGG 365  
QY 249 GGAACACTCCACCGCTGCGCCGGAAGCTGAGAGAGCTTCCCGAGCTTC----- 300  
DB 366 CCGGCGGTGACAGAGAGGAGCTCGAGAGGTGTAAGAGAGGAGGAGGCGCATGCTCCACG 425  
QY 301 -----GCCGAGCTTCGGGGGCTTCTGTCTCGGGGCTTACACCGGCGGCGGT 350  
DB 426 TACAGCAGAGACCTTCGACAGACTCTGCTTGGGCTGGGCGCTACACAGCTGGGCGCAT 485  
QY 351 GGCCTTCATGCTTCCTGAGAGCGGAGTGGCGGCGGTGACGAGAACTTCGAGGAGTCT 410  
DB 486 TGCCTTATGCGCTTTGGCGCAGCAACGCGTGTGGATGGCAACGATGACGCGGTCT 545  
QY 411 CTCGCGCTCTTCGCG 426

Db 546 GTGCCGTGTCCGAGCC 561

RESULT 9  
BI818560 771 bp mRNA linear EST 04-OCT-2001  
LOCUS 603033084F1 NIH\_MGC\_115 Homo sapiens CDNA clone IMAGE:5174084 5',  
DEFINITION mRNA sequence.  
ACCESSION BI818560  
VERSION BI818560.1 GI:15929676  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
1 (bases 1 to 771)  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgabbs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.jnl.gov  
Plate: LLM11433 row: n column: 21  
High quality sequence stop: 757.

FEATURES  
source  
1. 771  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5174084"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC 115"  
/note="Organ: pooled brain, lung, testis; Vector:  
PCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA  
source anonymous pool of 6 male brains, age range 23-27; 1  
male lung, age 27; and 1 male testis, age 69. Library is  
oligo-dT primed and directionally cloned (EcoRV site is  
destroyed upon cloning). Average insert size 1.8 kb.  
Insert size range 1-3 kb. Library is normalized and  
enriched for full-length clones and was constructed by C.  
Gruber (Invitrogen). Research Genetics tracking code  
021. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 9.2%; Score 90; DB 12; Length 771;  
Best Local Similarity 55.3%; Pred. No. 0.00065;  
Matches 208; Conservative 0; Mismatches 150; Indels 18; Gaps 1;  
QY 69 GGAGAGACCTTACCGCGCTCTGCTCCGAGGCTCTTTCGACAGACCCGGGTGGA 128  
DB 253 GGACAGCGGGCATATGCTGTGTGAGGTCTGACAGGTCATGCTGACAGACCCAGGTTGC 312  
QY 129 GCAGGCGCTCCCTATTACCGCGCTTCTGAGAGCGCTTTCGACAGGCGGCGCTTGC 188  
DB 313 CACTGTGATCACTACTATACCGATGATGAGAGAGGCTTACACTGACAGGACCTTGCC 372  
QY 189 CGCGGCTTCTGAGAGAGTCTTCAGGCTTCGACAGGCGGCGCTACTACCGGCGGC 248  
DB 373 CAGTGTCTCCCTGAGAGAGGTGAATCACTCTGGGCTGGCGCTTACTATTCGTGG 432  
QY 249 GGAACACTCCACCGCTGCGCCGGAAGCTGAGAGAGCTTCCCGAGCTTC----- 300  
DB 433 CCGGCGGCTGACAGAGAGGAGCTGGAAGGTGTGAGAGAGCTGAGGGGCGACATGCGACG 492  
QY 301 -----GCCGAGCTTCGGGGGCTTCTGTCTTCGGGCTTACACCGGCGGCGGT 350  
DB 493 TACAGCAGAGACCTTCGACAGACTCTGCTTGGCGCTGAGGCGGCTTACACAGCTGGGCGCAT 552

QY	351	GGCCCTCATGCCCTTGGGGAGCGGGTGCAGGAGACCTCCGAGGCTCT	410
Db	553	TGCCTTATGCGCTTTGGCCAGGCAACCGGTGTGTGATGACCACTAGCAAGGCTCT	612
QY	411	CTCCGCGCTCTTCCG 426	
Db	613	GTGCGGTCTCCGAGCC 628	
RESULT 10			
LOCUS	BX401779	1014 bp	mRNA linear EST 13-MAY-2003
DEFINITION	BX401779 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED		
ACCESSION	BX401779		
VERSION	BX401779.1	GI:30630494	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Bukacynski, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. (bases 1 to 1014)		
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Genoscope - Centre National de Sequencage Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 922.r For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSDDL008BD01Q1p1cluster=922.r. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CSDDL008BD01Q1p1.		
FEATURES			
SOURCE	Location/Qualifiers 1..1014 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CSDDL008YG02" /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED" /cell_line="RAMOS CELL LINE" /clone_id="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED" /note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."		
ORIGIN			
Query Match	9.2%;	Score 90;	DB 13; Length 1014;
Best Local Similarity	55.3%;	Pred. No. 0.00069;	
Matches	208;	Conservative 0;	Mismatches 150; Indels 18; Gaps 1;
QY	69	GGAGAGAGACCTTACCGCGTCTGTGTCGAGGTCTCTTGACAGACCCGGTGA	128
Db	487	GGACGAGGCGGCAATGCTGTGTGGGTCTCAGAGGTCTATGCTGACGAGACCCAGTTC	546
QY	129	GCAAGCCTCCCTATTACCGCGCTTCTGTGAGGCGCTTCCACCTGAAAGCCTTGGC	188
Db	547	CACGTGATCAACTACTATACCGATGAGTATCAGAAAGTGGCTTAACATGACGAGACTGGC	606
QY	189	CGCGGCTTCCCTGGAAGAGTCTTATAGGTCTGCGAGGCGGCGGCTACTACCGCGGAC	248
Db	607	CAGTGTCTCCCTGAGAGAGTGAATCAACTCTGGCGCTGGCCCTGGGCTACTATTCTCGGG	666
QY	249	GGAACTCTCAACGCGCTGGCCCGAAGCTGAGAGAGCTTCCCGGAGCTTC-----	300
Db	667	CCGGCGGTGAGAGAGGAGCTCGGAAGTGTATGAGAGAGCTAGAGGGCCACATGCCAG	726

QY	301	-----GCCAGCCTTCCGGGGGCTTCCTGGTGTGGGGGCTTACACGGGGCGCGGT	350
DB	727	TACAGCAGAGACCTCGACGACGCTTCCTCGGCTGGGGGCGCTACACGCTGGGGCCAT	786
QY	351	GGCCTTCATCCCTTTCCGGGAGCGGGGTGGCGCGGTGACCGGAGACGTCGAGGGGTCT	410
DB	787	TGCTCTTATGCCCTTTGGCCAGGACACCGGATGTGGATGGACAACTGATGACCGGGTCT	846
QY	411	CTCCGCTCTTGGCC	426
DB	847	GTGCGTGTCCGAGCC	862
RESULT 11			
AK077546	1933 bp	mRNA	linear
LOCUS			
DEFINITION	Mus musculus 8 days embryo whole body cDNA, RIKEN full-length enriched library, clone:573044C03 product:similar to ADENINE-DNA GYCOSYLASE. [Mus musculus], full insert sequence.		
ACCESSION	AK077546		
VERSION	AK077546.1	GI:26097402	
KEYWORDS	HTC, CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1		
TITLE	Carninci, P. and Hayashizaki, Y.		
JOURNAL	High-efficiency full-length cDNA cloning		
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)		
PUBMED	99279253		
REFERENCE	10349686		
AUTHORS	2		
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
MEDLINE	Genome Res. 10 (10), 1617-1630 (2000)		
PUBMED	20499374		
REFERENCE	11042159		
AUTHORS	3		
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsumata, T., Tashiro, H., Itoh, M., Suni, N., Ishi, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Watanabe, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Iwata, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsui, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
JOURNAL	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)		
PUBMED	20530913		
REFERENCE	11076861		
AUTHORS	4		
TITLE	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
JOURNAL	Functional annotation of a full-length mouse cDNA collection		
REFERENCE	Nature 409, 685-690 (2001)		
AUTHORS	5		
TITLE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.		
JOURNAL	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		
REFERENCE	Nature 420, 563-573 (2002)		
AUTHORS	6 (bases 1 to 1933)		
TITLE	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Horikawa, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imochi, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numaoka, R., Ohno, M., Ohashi, N.,		



QY 249 GGAACACCTCCAGCCCTGGCCCGAAGCGTGAGAGCTTCCCGAGCTTC----- 300  
 Db 501 CCGGCGGCTGAGAGAGGAGCTCGAAGGTGTAGAGAGCTAGGGGGCCACATGCCACG 560  
 QY 301 -----GCCGAGCTTGGGGGGCTTCTGTCTCGAGCTTACACCGCGCGCGGT 350  
 Db 561 TACAGACAGAGACCTCGACAGCTCTGCTGGCGGTGGCGGTACACAGCTGGGGCAT 620  
 QY 351 GGCTTCATCGCTTCCTGGGGA-GCGGCTGGCGCGGTGGAAGGAGCTCGAGAGGTCC 409  
 Db 621 TGCCTATATCCCTTTGGCCAGGCAACCGGTGTGTGATGGCAACCTAGACCGGGTGC 680  
 QY 410 TCTCCGCTCTTGGGCC---GGGAAAGCCCCAAGAGAGAGAGCTTTCGCGCTCGCC 466  
 Db 681 TGTGCTGTGTCGAGGACCTGTGTCTATCCAGACAGACCTTGTTCCTCAGACGTCT 740  
 QY 467 AGGACCTCTCTCCCGAGGCGGTGAC-----CGGGGTGTGGAACCAAGCCCTCAT 518  
 Db 741 GGGGTCTAGCCAGAGAGCTGTGTGACCAACCGGGCCAGGAATTCACCAAGAGCGCAT 800  
 QY 519 GGAGCTCGGGCCACGCTGTGCTGCGAAGCGCCCTTGGCGGGCTTGCCTCTAG 577  
 Db 801 GGAGCTAGGGGCCCGAGTGTGTACCCCAACGCGCCACTGTGACAGCGGCGCTGTGG 859

RESULT 13  
 Bg761186 900 bp mRNA linear EST 15-MAY-2001  
 LOCUS 602718527F1 NIH\_MGC\_49 Homo sapiens cDNA clone IMAGE:4858374 5',  
 DEFINITION mRNA sequence.  
 Bg761186  
 VERSION Bg761186.1 GI:14071839  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 900)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@b-remail.nih.gov  
 Tissue Procurement: ATCC/DCTD/DTM  
 DNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: L1CM1712 row: F column: 07  
 High quality sequence stop: 867.

## FEATURES

source

1. 900  
 location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4858374"  
 /tissue\_type="melanotic melanoma, high MDR (cell line)"  
 /lab\_host="RDH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_49"  
 /note="Organ: skin; Vector: pOTB7; Site: 1: XhoI; Site: 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCGAGG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 library."

ORIGIN

Query Match 9.0%; Score 87.4; DB 12; Length 900;  
 Best Local Similarity 54.9%; Pred. No. 0.0016;  
 Matches 207; Conservative 0; Mismatches 151; Indels 19; Gaps 1;  
 QY 69 GGAAGAGACCTTACCGCGTCTGTCTCCGAGTCTTCTGACAGACCCGGGTGA 128  
 Db 331 GGACAGGCGGGCATATGCTGTGTGTGTCTCAGAGGTCTGACGACGACCCAGGTGC 390  
 QY 129 GCAGGCCCTCCCTATTATACGCCCGCTTCTGAGGCTTTCACCCAGAGGCCCTGGC 188  
 Db 391 CACTGTATCAACTACTATACCGATGATGATGATGATGATGATGATGATGATGATGAT 450  
 QY 189 GCGGCTTCCCTGGAAGAGTCTTGAAGTCTGCGAGGGCGGGCTACTACCGCGGAC 248  
 Db 451 CAGTCTTCCCTGGAAGAGTGTATCACTCTGGGCTGGCCCTGAGCTATTCCTGCG 510  
 QY 249 GGAACACTTCACCGCGCTGCGCCCAAGAGCTGAGAGCTTCCCGAGCTTC----- 300  
 Db 511 CCGGCGGCTGACAGAGGAGGAGCTCGAAGGTGTAGAGAGCTAGGGGCCACATGCCACG 570  
 QY 301 -----GCCGAGCTTGGGGGCTTCTGTGCTTGGGCTTACACCGCGCGCGG 349  
 Db 571 TACAGACAGACCTTGCAGAGCTCTGCTGCGGTGGGCGCTACACAGCTGGGGCA 630  
 QY 350 TGGCTCCATGCTTGGGAGCGGAGCGGCTGCGAGCGGAACTCCGAGAGGTCC 409  
 Db 631 TTGCTCTATGCTCTTGGCCAGGCAACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 690  
 QY 410 TCTCCGCTCTTGGCC 426  
 Db 691 TGTGCGGTGTCGAGCC 707

RESULT 14  
 Bx415111/c 982 bp mRNA linear EST 15-MAY-2003  
 LOCUS Bx415111 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP004Y120  
 DEFINITION 5-PRIME, mRNA sequence.  
 Bx415111  
 VERSION Bx415111.1 GI:30765470  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 982)  
 Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. Contact : Feng Liang Email : fliang@life-tech.com URL :  
 http://fulllength.invitrogen.com/InvitrogenCorporation1600  
 Faraday Avenue Genoscope sequence ID : CS0CAP004B10QPL.  
 location/Qualifiers  
 1. 982  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0CAP004Y120"  
 /tissue\_type="THYMUS"  
 /clone\_lib="Homo sapiens THYMUS"  
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
 with a NotI-oligo(dT) primer. Five prime end enriched,  
 double-strand cDNA was digested with Not I and cloned into  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 Library was not normalized."

## FEATURES

source

1. 982  
 location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0CAP004Y120"  
 /tissue\_type="THYMUS"  
 /clone\_lib="Homo sapiens THYMUS"  
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
 with a NotI-oligo(dT) primer. Five prime end enriched,  
 double-strand cDNA was digested with Not I and cloned into  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 Library was not normalized."

ORIGIN

Query Match 8.8%; Score 85.8; DB 13; Length 982;







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 4, 2004, 05:07:23 ; Search time 1258 Seconds  
(without alignments)  
2834.655 Million cell updates/sec

Title: US-09-938-901-1

Perfect score: 975  
Sequence: 1 gtgagagcccttcgaggaagc.....tagtcccccccgagca 975

Scoring table: IDENTITY NUC  
Gap 10.0, Gapext 1.0

Searched: 2421054 seqs, 1828716029 residues

Total number of hits satisfying chosen parameters: 4842108

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	975	100.0	975	10 US-09-938-901-1	Sequence 1, Appl1
2	112.4	11.5	939	14 US-10-156-761-4691	Sequence 4691, Ap
3	112.4	11.5	9025608	14 US-10-156-761-1	Sequence 1, Appl1
4	104	10.7	1878	9 US-09-925-301-484	Sequence 484, App
5	92.8	9.5	528	9 US-09-974-300-1583	Sequence 1583, Ap
6	89.2	9.1	1660	12 US-10-425-114-32132	Sequence 32132, A
7	86.6	8.9	912	14 US-10-216-817-31	Sequence 31, Appl1
8	86.6	8.9	1312	14 US-10-216-817-21	Sequence 21, Appl1
9	82	8.4	9909	12 US-10-158-844-12	Sequence 12, Appl1
10	81.2	8.3	968	9 US-09-864-866-46	Sequence 46, Appl1
11	71.4	7.3	930	14 US-10-156-761-4577	Sequence 4577, Ap
12	71.4	7.3	9025608	14 US-10-156-761-1	Sequence 1, Appl1
13	68.2	7.0	991	15 US-10-398-221-1588	Sequence 1588, Ap
14	68.2	7.0	3155	15 US-10-398-221-3596	Sequence 3596, Ap
15	66.8	6.9	11058	14 US-10-156-761-3629	Sequence 3629, Ap

16	66.8	6.9	18876	14 US-10-329-079-42	Sequence 42, Appl1
17	66.8	6.9	61944	14 US-10-329-079-34	Sequence 34, Appl1
18	64.6	6.6	1230	14 US-10-156-761-5404	Sequence 5404, Ap
19	63.8	6.5	8244	15 US-10-402-842-3	Sequence 3, Appl1
20	63.8	6.5	47988	15 US-10-402-842-1	Sequence 1, Appl1
21	62.4	6.4	32768	9 US-09-070-927A-17	Sequence 17, Appl1
22	62.2	6.4	1227	15 US-10-458-201-23	Sequence 23, Appl1
23	62.2	6.4	15120	15 US-10-458-201-1	Sequence 1, Appl1
24	61.4	6.3	1230025	15 US-10-289-762-1	Sequence 1, Appl1
25	61.2	6.3	975	10 US-09-938-901-1	Sequence 1, Appl1
26	60.4	6.2	566	12 US-10-424-599-105239	Sequence 105239,
27	59.8	6.1	10325	10 US-09-938-901-5	Sequence 5, Appl1
28	59.8	6.1	5355	14 US-10-205-032-19	Sequence 1, Appl1
29	59.8	6.1	60196	14 US-10-205-032-1	Sequence 1, Appl1
30	58.4	6.1	272	11 US-09-864-408A-9041	Sequence 9041, Ap
31	58.4	6.0	1122	14 US-10-156-761-4667	Sequence 4667, Ap
32	58.8	6.0	1128	12 US-10-671-403-106	Sequence 106, App
33	58.8	6.0	1128	12 US-10-671-419-106	Sequence 106, App
34	58.8	6.0	1365	14 US-10-156-761-6400	Sequence 6400, Ap
35	57.6	5.9	4437	14 US-10-156-761-4189	Sequence 4189, Ap
36	57.4	5.9	1404	14 US-10-156-761-2469	Sequence 2469, Ap
37	57.4	5.9	1590	12 US-10-671-403-3	Sequence 3, Appl1
38	57.4	5.9	1590	12 US-10-671-419-3	Sequence 3, Appl1
39	57.4	5.9	2007	12 US-10-671-403-1	Sequence 1, Appl1
40	57.4	5.9	2007	12 US-10-671-419-1	Sequence 1, Appl1
41	57.4	5.9	2934	10 US-09-938-901-7	Sequence 7, Appl1
42	57.2	5.9	1728	14 US-10-156-761-3783	Sequence 3783, Ap
43	57.2	5.9	319630	15 US-10-398-221-7	Sequence 7, Appl1
44	57.2	5.9	3011208	15 US-10-398-221-2058	Sequence 2058, Ap
45	57	5.8	11238	14 US-10-205-032-15	Sequence 15, Appl1

## ALIGNMENTS

RESULT 1					
US-09-938-901-1					
Sequence 1, Application US/09938901					
Publication No. US20030008291A1					
GENERAL INFORMATION:					
APPLICANT: Kuramitsu Seiki,					
APPLICANT: Yokoyama Shigeyuki					
TITLE OF INVENTION: GENE ENCODING DNA REPAIR ENZYME					
FILE REFERENCE: PH-1261-US					
CURRENT APPLICATION NUMBER: US/09/938, 901					
CURRENT FILING DATE: 2001-08-24					
PRIOR APPLICATION NUMBER: JP2001-47762					
PRIOR FILING DATE: 2001-02-23					
NUMBER OF SEQ ID NOS: 17					
SOFTWARE: PatentIn Ver. 2.0					
SEQ ID NO 1					
LENGTH: 975					
TYPE: DNA					
ORGANISM: Thermus thermophilus					
FEATURE:					
NAME/KEY: CDS					
LOCATION: (1)..(975)					
US-09-938-901-1					
Query Match					
Best Local Similarity 100.0%; Score 975; DB 10; Length 975;					
Matches 975; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	GTGAGGCGCTTCGAGGAAGCCCTCTCGCTGATACCGGGAAGAGCCGCCCTCTCC	60		
DB	1	GTGAGGCGCTTCGAGGAAGCCCTCTCGCTGATACCGGGAAGAGCCGCCCTCTCC	60		
QY	61	TGCGGGGGGAGAGACCTTACCGCGCTCTGTCTTCGAGGTCCTTCTGACAGACC	120		
DB	61	TGCGGGGGGAGAGACCTTACCGCGCTCTGTCTTCGAGGTCCTTCTGACAGACC	120		
QY	121	CGGGTGGAGCAGGCGCTTCTTATACCGCGCTTCTGAGGCGCTTCCACCTGAAG	180		

```

Db      121 CGGGTGAAGACAGCCCTCCCTATTACCGCCGCTTTCTGAGCGCTTCCCACTGAAG 180
Qy      181 GCCCTGACCGGCTTCTCTGAAAGAGTCTTAGGGTCTGCGAGGGGCGGGCTACTAC 240
Db      181 GCCCTGACCGGCTTCTCTGAAAGAGTCTTAGGGTCTGCGAGGGGCGGGCTACTAC 240
Qy      241 CGGGGAGGAGAACCTTCCACCGCTGCGCCGGAAGCGTGAAGAGCTTCCCGGAGCTTC 300
Db      241 CGGGGAGGAGAACCTTCCACCGCTGCGCCGGAAGCGTGAAGAGCTTCCCGGAGCTTC 300
Qy      301 GCCGAGCTTGGGGGCTTCTGCTCTGAGCTTACACCGCGCGGCTTGGCTTCCATC 360
Db      301 GCCGAGCTTGGGGGCTTCTGCTCTGAGCTTACACCGCGCGGCTTGGCTTCCATC 360
Qy      361 GCCTTGGGGAGCGGGTGGCGGCGGTGAACGAGAACCTTCGAGAGGCTCTTCCCGCTTC 420
Db      361 GCCTTGGGGAGCGGGTGGCGGCGGTGAACGAGAACCTTCGAGAGGCTCTTCCCGCTTC 420
Qy      421 TTGCGCCGGGAAAGCCCAAGAGAGAGCTTTTCCGCTTCCGAGGCTCTCTCC 480
Db      421 TTGCGCCGGGAAAGCCCAAGAGAGAGCTTTTCCGCTTCCGAGGCTCTCTCC 480
Qy      481 GAGGGGTGAACCCCGGGGTGTGAACCAAGGCTCATGAGCTTGGGCGACAGCTTCC 540
Db      481 GAGGGGTGAACCCCGGGGTGTGAACCAAGGCTCATGAGCTTGGGCGACAGCTTCC 540
Qy      541 CTGCGGAAACGCGCCGTTGCGGGGCTGCGCCCTTGAAGGCTTCTGCGGGGAGAG 600
Db      541 CTGCGGAAACGCGCCGTTGCGGGGCTGCGCCCTTGAAGGCTTCTGCGGGGAGAG 600
Qy      601 GCGCCGCGGCGCTACCCGCGCCAGAGAGCGCGGAGAGAGAGAGAGAGAGAGAG 660
Db      601 GCGCCGCGGCGCTACCCGCGCCAGAGAGCGCGGAGAGAGAGAGAGAGAGAGAG 660
Qy      661 CTGCTCTCTCTGAGGCGGAGAGGGGTGACCTGAGAAAGCTTGAAGGGGCTTCAAGGC 720
Db      661 CTGCTCTCTCTGAGGCGGAGAGGGGTGACCTGAGAAAGCTTGAAGGGGCTTCAAGGC 720
Qy      721 CTCTACGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
Db      721 CTCTACGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
Qy      781 GTGAGGTCTAGAGCCCTTAGAGAGAGTGCACGCTTCAACCAACGAGAGCTTCCGCTG 840
Db      781 GTGAGGTCTAGAGCCCTTAGAGAGAGTGCACGCTTCAACCAACGAGAGCTTCCGCTG 840
Qy      841 GAGGTGCGGGGGGCTTTGGAGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db      841 GAGGTGCGGGGGGCTTTGGAGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Qy      901 AAGCTATGAGAGAGTGTCTCGCAAGGCGCTTCTCTCTCTCTCTCTCTCTCTCTCT 960
Db      901 AAGCTATGAGAGAGTGTCTCGCAAGGCGCTTCTCTCTCTCTCTCTCTCTCTCTCT 960
Qy      961 CCCCTCCCGAGCGCA 975
Db      961 CCCCTCCCGAGCGCA 975

```

# RESULT 2

```

; Sequence 4691, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMIURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262

```

```

; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; SEQ ID NO 4691
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(939)
US-10-156-761-4691

```

```

Query Match      11.5%; Score 112.4; DB 14; Length 939;
Best Local Similarity 52.1%; Pred. No. 3e-15; Mismatches 306; Indels 33; Gaps 4;
Matches 369; Conservative 0;

```

Qy	4	GAGGCTGCGGAAAGCCCTCTGCTGTAACGAGAAAGCCCGCCCTCCCTG	63
Db	73	GAGGCTGCAAGCCAGTCAATGATTCAGACAGACGCGCTGACCTCCGTG	132
Qy	64	CGGG---GAGAAAGACCTTACCGGCTCTGCTCCGAGTCTTTCAGACAGACC	120
Db	133	CGCGCCCGGAGCGCGGCTGAGTGTCAATGATTCAGACAGACGCGCTGAC	192
Qy	121	CGGCTGAGAGAGCCCTTCTCTTATTAACCGCGCTTCTGAGCGCTTCCACCTGAAG	180
Db	193	CGGCTCAACCGCGTGTGCTGCTCTTACAGACAGTGTGCGCGCTGAGCGCGCTG	252
Qy	181	GCCCTGCGCGGCTTCTCTGAGAGAGTCTTGAAGGTCTGAGAGGGGCGCTTACTAC	240
Db	253	GACCTCCGCAAGAGAGGCGCGCGAGAGCGTCTGCGCTGAGGCGCGCTGACCC	312
Qy	241	CGGCGGCGGAAACCTTCAACCGCTTGGCC-----GAGCGTGAAGAG	285
Db	313	CGGCGCGCTGCGCTGCAAGCGCGCGCTGATTAACGAAACGCAACGCGAG	372
Qy	286	CTTCCCGAGCTTGGCGAGCTTGGGGGCTTCTGCTGCTGAGGCTTACCGCGG	345
Db	373	GTAACCAAGAGAGAGCGCGCTGCTGCGCTGCGCGATGAGGAAATACACCGCG	432
Qy	346	GCGGTGCTTCAATGCTTCTGAGAGCGGAGCGGCGGCTGAGAGAGAGAGAGAG	405
Db	433	GCGTGCCTGCTTCTGCTTACGCGCGAGCGCACGCGTCTTGAACCAAGTCCGCT	492
Qy	406	GTCCTTCCCGGCTTCTGCGCGGAAAGCCCAAGAGAGAGAGAGAGAGAGAGAG	459
Db	493	GTCCTTCCCGGCTTCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	552
Qy	460	-----CTGCGCGAGGCTCTCTCCGAGGCGTGAACCGGGG---TGTGAACAG	510
Db	553	CGGAGCTGCGCGCGGCTGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	612
Qy	511	GCCCTAGAGAGTCTGAGGCGACAGTCTGCTGCGGAAACGCGCGCTTGGGGCTG	570
Db	613	GCTCTATGAGAGTCTGAGGCGGCTGCTGCAAGGAGAGAGAGAGAGAGAGAGAG	672
Qy	571	CCCTAGAGGCTTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	630
Db	673	CGATGCGCGGCGAGTGTGCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	732
Qy	631	CGCGGAG	678
Db	733	CGCGGAG	780

## RESULT 3

```

; Sequence 1, Application US/10156761
; Publication No. US20030119018A1

```



OY 579 GGCCCTTCGCGGG 592  
Db 1049 GAGCCTGTGCGGG 1062

## RESULT 5

US-09-974-300-1583  
; Sequence 1583, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Ib Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: Expression  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; PRIOR FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1583  
; LENGTH: 528  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(528)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-974-300-1583

Query Match 9.5%: Score 92.8; DB 9; Length 528;  
Best Local Similarity 53.4%; Pred. No. 5.1e-11;  
Matches 227; Conservative 0; Mismatches 183; Indels 15; Gaps 1;

OY 17 AAGCCCTCTCGCTGTACCGGGAAACGCGCCCTCCCTCGGGGGAGAAAG 76  
Db 43 AAGATTAATTCTGTGTATGACCAAGAAAAGCGGATCGCTTGCGTACAGCG 102  
OY 77 ACCCTTACCGCGCTCGGTCTCCGAGGTCTTTCGACAGACCCGGGTGAGCAGGCC 136  
Db 103 ACCCTTAATAGGTGTGGGTGTCCGAAGTATGCTTACAGACGAGAGTACACGCTGA 162  
OY 137 TCCCTATTACCGCGCTTCTGTGAGGCTTCCCAACCGTGAAGGCGCTGGCGCGCTT 196  
Db 163 TTCTTACTTCAACAATTATGAAAAGTTTCCGACCTGGAGCGCTGCTGAAGCG 222  
OY 197 CCTGGAAGAGTCTTAAAGGTCTGACAGGGCGGCTACTTACCGCGGGCGAAACAC 256  
Db 223 ACGAAGAAAAGTCTTAAAGCTGGGAAGACTGGGATATTAATTCAGAGGTTCCGAAC 282  
OY 257 TCCAGCGC-----CTGGCCCGAAGCGTGAAGAGCTTCCCGAGCTTCG 301  
Db 283 TGCAGAGCGCTGTACAGGAAGTTACGAACGATACGAGGTGTGTCTCTTCAAAAG 342  
OY 302 CCGAGCTTGGGGGCTTCTGTGTCTGCGCTTACACCGCGGCGGCGGCTCCATCG 361  
Db 343 AAGAAATTGGGAGCTTAAAGCGGTGCGCTTATACAGAGGCGGCGTGTGAGCATCG 402  
OY 362 CCTTGGGAGAGGGGTGGCGGCGGTGAGCGGAACGTTCGAGGGTCTCTCCCGCTCT 421  
Db 403 CCTACATACGCGGTGCGGCGGTGAGCGGAATGTCAATGGGGTCAATGTACGGAATTC 462  
OY 422 TCGCC 426  
Db 463 TTTCC 467

RESULT 6  
US-10-425-114-32132  
; Sequence 32132, Application US/10425114

; Publication No. US2004003488B1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 32132  
; LENGTH: 1660  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMFLB73265D07\_F1  
US-10-425-114-32132

Query Match 9.1%: Score 89.2; DB 12; Length 1660;  
Best Local Similarity 50.7%; Pred. No. 2.4e-10;  
Matches 282; Conservative 0; Mismatches 253; Indels 21; Gaps 2;

OY 60 CTGCGGGGAGAGAACCTTAACCGGCTCTGTCTCCAGGTCTCTTCCAGCAGAC 119  
Db 278 CGGACGAGAGAGAGAGGCGTACGCGGTGTGGGTGCGAGGTATGCTGACGAGAC 337  
OY 120 CCGGTGAGAGAGCGCTCCCTTATACCGCCCTTCTGAGAGCTTTCCACCTCGAA 179  
Db 338 GGGGTGCGCGTGTCTGCTTCTTCTACAGAGGTGATGCGGCGGCGACCTGACG 397  
OY 180 GACCTGCGCGGCTTCTCTGAAGAGTCTTACGAGGTCTGCGAGGGCGGCTACTA 239  
Db 398 AAGCTGCGCGCGCTTACGAGAGAGTGAAGATGTGGCGGCTCTTGCTACTA 457  
OY 240 CCGCGGCGGAGAACCTTC-----ACCGCTGCGCGGAGAGCTGAGAGCTTCC 290  
Db 458 CCGAGGCTCTGATTTCTGTGAGAGAGCAAGCAATATCGAAAAGGGTGTTC 517  
OY 291 CCGAGCTTGCAGGCTTGGGGGCTTCTGTGTCTGCGGCTTACACCGCGCGCGGT 350  
Db 518 TTGACGCGCTTACCTTGTGTGAGTTCTGTGATTTGGGATTTACACCTGAGCAAT 577  
OY 351 GACCTCATGCTTTCGGGAGCGGCTGCGGCGGTGACGGGAACGTCCGAGGCTCT 410  
Db 578 CGTTCAATTGCTTCAACGAGGTGTCCAGTTGTGATGAAATGTATACAGATCAT 637  
OY 411 CTCCGCTCTTGTGCGCGGGAAGCCCAAGAGAA-----GAGCTTTTCCG 458  
Db 638 CAGAGGCTTAAACCAATGCTGACAAACCAAGGAATCTCAACAGTGAAGATTCG 697  
OY 459 CTTGCGCGAGGCTCTTCCCGAGGCGGTGACCGGGGGTGTGAACAGGCGCTCAT 518  
Db 698 GACCTGTAGTCAATGTTGATACCTTTGAAACAGAGATTTAAACCAAGCATGAT 757  
OY 519 GAGACTGCGGCGCAAGCTGTGCTGCGGAACCGCCGTTGGCGGCGCTGCGCCCTAG 578  
Db 758 GAGTTAGAGCAAACTTATGTAGCAAAAGCAAGCTGTGTCTCAATGCCAATTC 817  
OY 579 GGCCTTGTGCGGGGG 594  
Db 818 TAGCACTGCCAAGCG 833

RESULT 7  
US-10-216-817-31  
; Sequence 31, Application US/10216817  
; Publication No. US20030129619A1  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE

```

: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
: TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
: TITLE OF INVENTION: IN GENES OF THE MTC1 FAMILY
: FILE REFERENCE: 03495.0233-00000
: CURRENT APPLICATION NUMBER: US/10/216,817
: CURRENT FILING DATE: 2002-11-13
: PRIOR APPLICATION NUMBER: 60/311,824
: PRIOR FILING DATE: 2001-08-14
: PRIOR APPLICATION NUMBER: 60/313,523
: PRIOR FILING DATE: 2001-08-21
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 31
: LENGTH: 912
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: US-10-216-817-31

```

Query Match	8.9%	Score	86.6	DB	14	Length	912
Best Local Similarity	50.2%	Pred. No.	9.6e-10				
Matches	345	Conservative	0	Mismatches	309	Indels	33
						Gaps	4

0y	22	CTCTCGCCTGATACCGGAAAAAGCCCCCGCCCTCCCTCGGCGGGG---	GAGAAGGAC	78
Db	61	CTTCTCGCTTGGATACGCGATCGCACCGGGGATCTGCCCTGGCGAGAGCCCGGTGCAGC		120

0y CCTTACCGGCTCTGTCTCCGAGGCTCTTCTGACGACAGCCCCGGGTGACAGCAGCCCTC 138

Db 121 CCGTGGCAGATCTCTGTACGCGAGTTTCACTGACGACAGACGCCGCCGCCGCCGGGTGCTG 180

oy 139 CCCATTACCGCCCGCTTTCGGAGCCGTTCCCAACCTGAAGCCCTGGCGCGGCTTCC 198  
 Db 181 GCCATCTGCGCGAGCTGGGTCCGCGCGTGGCCCAAGCCCGTGGCCACCGCCACGCCACG 240

0y CTGGAAGAGCTCTTAGGGCTCTGGCAGGGGGGGGGCTACACCGCGGGGGAACACTC 258  
 Db 241 ACCGCGGAGTGTACGCGGCTGGGGCAGCTGGCTATCCAGCGAGCGCAACGCTTAA 300

07 CACCGCCCTGCCCCGAGCGTG-----GAGGAGCTTCCCGGAGCTTGGCC 303  
259 CACCGCCCTGCCCCGAGCGTG-----GAGGAGCTTCCCGGAGCTTGGCC 303  
301 CACGAGTGGCCACCGTCATGCGCCGACCAACAATGACGTGTGCCCGACGATATCGAG 360

308 GAGGCTCGGGGCTTCCTGGCTTCCTGCGCCCTTACATCGCGCGCGGAGTGGTCCATCGCC 363  
 361 ATCTGTGTCACCTGCGCGGCGTTCGGGAGCTACACCGCGCGCGGTGGCGTGTTCGCT 420

421 TACCGCCAGCGGTCGGGTGTGACACCAATGTGCGCGCGTGTGGCCGCCGCGCTT 480

Dc  
481 CACGCGCCGCCGAGCGCCGATGCATCGGTGGCGCGCACCACGCCGACGTCTTGGCG 540

Dc  
Db

541 CTGTGCGGCACCGCAGACGGCCCTGAATTTTCGTGCGCGCTGATGGAGTTGGGTGCG 600

Db 601 ACCGATGTCACCCGCCGACACCCCGGTCGGTTATGCCCGCTGAGACTGCTGCGCATGG 660

Db 661 CGGCATGCCGGTTATCCGCCCTCGACGGTCC --- GCCCGCGCGGGGGGCGAGGCTTACACC 717

Db

718 GGAACCGACCGCCAAAGTCCCGCGACGG 744

RESULT 8  
US-10-216-817-21

```

: Sequence 21 Application US/10216817
: Publication No. US20030128619A1
: GENERAL INFORMATION:
: APPLICANT: GICQUEL, BRIGITTE
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
: TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
: TITLE OF INVENTION: IN GENES OF THE MULT FAMILY
: FILE REFERENCE: 03495,0233-00000
: CURRENT APPLICATION NUMBER: US//10/216,817
: CURRENT FILING DATE: 2002-11-13
: PRIOR APPLICATION NUMBER: 60/311,824
: PRIOR FILING DATE: 2001-08-14
: PRIOR APPLICATION NUMBER: 60/313,523
: PRIOR FILING DATE: 2001-08-21
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 21
: LENGTH: 1312
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: US-10-216-817-21

```

```

Query Match      8.9%;      Score 86.6;  DB 14;      Length 1312;
Best Local Similarity 50.2%;  Pred. No. 8, 9e-10;
Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

```

0y 22 CTCTCGCTGTGATCCGGAAAAAGCCGCCCTCTCCCTGTGGGGGG---GAGAGAGAC 78  
 Db 261 CTCTCGCTGTGATCAACGATCCGACCCGGGATCTGCCCTGGGAGAGACCCGGTGTACAC 320

0y 79 CCTTACCCCGTCCGTGTCCTCGAGGTCCTTCTCAGCAGACCCGGGTGAGCAGGCCCTC 138  
 Db 321 CCTGGGAGATCCTGTGTGTCAGCGATTATGCTGACAGAGACGCGCGCGCCCGGGTGTG 380

381 GCGATCTGCGCCGGAAGTGGGTGCGCGCGGTGGCCAGCGCGTGGCCACCGCCACGGCCAGC 440

Oy CTTGGAAGAGGCTCCCTTAGGGCTCTCTGGCAGGGGGCCCGGCTACCTACCTGGCCGGCGGATCACCCTC 258  
 Db 441 ACCGCGCATGTGTTACCGCGCTGGGGGCAAGCTGGGGCTATCCCGAGGGCGAGCCAAAGCGCTTAA 500

DB  
501 CACGAGTCCGCCCGCATTCGCCCGCACCACCAATGACGTGGTGCCCCGCAGCATATATCGAG 560

561 ATCTGTGTCACCCCTGCGGGCGTCGAGCTACACCGCGCGCGGTTCGCT 620

621 TACCGCCAGCGGGTCCGGTGTGTGAACCAATGTGCCGCGCGTGTGGCCCCGCCGTT 680

681 CACGGCCGCCCAAGCCCGGTGGCCCATTCGTGCGCGCGAGCCACCAAGCTTTGAGC 740

741 CTGTTCCGCA CCGCGAGACGCGCCCTAATTTCCGTCGCGCGTGATGAGATTGGCTGC 800

801 ACCGTGTCACCCGCCGACACCCCGGTATGCGGTGACCTGCTGCCGATGG 860

Db  
861 CGGCATGCGCGTTATCCGCCGTCGAGACGCTCC---GCCGGCGCGCGGGCAGGCGCTACACC 917

918 GGACCGACCGCCAGTCCGCGACGG 944

```

: Sequence 21 Application US/10216817
: Publication No. US20030128619A1
: GENERAL INFORMATION:
: APPLICANT: GICQUEL, BRIGITTE
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG
: TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS
: TITLE OF INVENTION: IN GENES OF THE MULT FAMILY
: FILE REFERENCE: 03495,0233-00000
: CURRENT APPLICATION NUMBER: US//10/216,817
: CURRENT FILING DATE: 2002-11-13
: PRIOR APPLICATION NUMBER: 60/311,824
: PRIOR FILING DATE: 2001-08-14
: PRIOR APPLICATION NUMBER: 60/313,523
: PRIOR FILING DATE: 2001-08-21
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 21
: LENGTH: 1312
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: US-10-216-817-21

```

```

Query Match      8.9%;      Score 86.6;  DB 14;      Length 1312;
Best Local Similarity 50.2%;  Pred. No. 8, 9e-10;
Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

```

0y 22 CTCTCGCTGTGATCCGGAAAAAGCCGCCCTCTCCCTGTGGGGGG---GAGAGAGAC 78  
 Db 261 CTCTCGCTGTGATCAACGATCCGACCCGGGATCTGCCCTGGGAGAGACCCGGTGTACAC 320

0y 79 CCTTACCCCGTCCGTGTCCTCGAGGTCCTTCTCAGCAGACCCGGGTGAGCAGGCCCTC 138  
 Db 321 CCTGGGAGATCCTGTGTGTCAGCGATTATGCTGACAGAGACGCGCGCGCCCGGGTGTG 380

381 GCGATCTGCGCCGGAAGTGGGTGCGCGCGGTGGCCAGCGCGTGGCCACCGCCACGGCCAGC 440

Oy CTTGGAAGAGGCTCCCTTAGGGCTCTCTGGCAGGGGGCCGGGCTACCTACCTGGCCGGCGGATCACCCTC 258  
 Db 441 ACCGCGCGATGTGTTACCGCGCGCTGGGGGCAAGCTGGGGCTATCCCGAGGGCGAGCCAAAGCGCTTAA 500

DB  
501 CACGAGTCCGCCCGCATTCGCCCGCACCACCAATGACGTGGTGCCCCGCAGCATATATCGAG 560

[illegible]

621 TACCGCCAGCGGGTCCGGTGTGTGAACCAATGTGCCGCGCGTGTGGCCCCGCCGTT 680

681 CACGGCCGCCCAAGCCCGGTGGCCCATTCGTGCGCGCGAGCCACCAAGCTTGGCG 740

741 CTGTTCCGCA CCGCGAGCGGCGCCCTAATTTTCGTCGCGCGTGATGAGATTGGGTCG 800

Dbb  
801 ACCGTGTCCACCCGCCGACACCCCGGTGCGGTTATCCCGCTGCACCTGCTGCCGATGG 860

Db  
861 CGGCATGCGCGTTATCCGCCGTCGAGACGGTCC---GCCGGCGCGCGGGCAGGCTACACC 917

918 GGACCGACCGCCAGTCCGCGACGG 944



RESULT 11  
US-10-156-761-4577  
; Sequence 4577, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 4577  
; LENGTH: 930  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(930)  
US-10-156-761-4577

Query Match 7.3%; Score 71.4; DB 14; Length 930;  
Best Local Similarity 47.1%; Pred. No. 1.7e-06;  
Matches 305; Conservative 0; Mismatches 321; Indels 21; Gaps 2;  
QY 39 GAAAAAGCCCGCCCTCCCTGCGGGGAGAGAGACCTTACCGCGTCTGCTC 98  
DB 207 GGTGTACCCGTACACACCCGAGTTGACTTGAGAACTCTTCCAGCTGATCTTGGC 266  
QY 99 CGAGGTCTTCTGAGAGAGACCGGGTGAAGCAGCCCTCCCTATTACCGCGCTTCT 158  
DB 267 GACCGTCTGTGCGCGAGACCACTGCGGTGAACCAAGACGCGCGCTGTT 326  
QY 159 GGAAGCGTTTCCACCTGAGAGCGCTGCGCGCTTCTTCCGAGAGAGTCTTAGAGT 218  
DB 327 CGCNAATACCGACCCCGAGAGACTGGCGCGCCCAACCCGAGAGGTGAGAGAT 386  
QY 219 CTGGCAGGGGGCGGCTACTACCGCGC-----GGCGAACACTTCCA 260  
DB 387 CTTCCGCGCCACCGGCTTCTCCGGGCCAAGACCAAGTCGTATAGGGTGTCCAAGGC 446  
QY 261 CCGCGTGGCCGAGAGCGTGAAGAGCTTCCCGAGCTGCGCAGCTTCCGGGGCTTCC 320  
DB 447 CTTGTGAGAGATTGCGCGCGAGGTCCCGGGCGCTCGAAGATCTCTCAAGCTGCC 506  
QY 321 TGTCTCGGCTTACACCGCGCGCGGTGAGCTTCCATCGCTTCCGGGAGCGGGTGGC 380  
DB 507 CGGGTGAAGAGCAAGAGCGGCTTCTGTGTCTGGGCAACCTTCCGGCGCGCGCAT 566  
QY 381 GCGCGTGAAGAGAAAGTCCGAGAGGTCTTCTCCGCTCTTCCCGGGAAGCCCCAA 440  
DB 567 CACCGTGAACACCATTTTCCAGCGGCTGTCCGCGTGAAGTGAACCGAGAGAGA 626  
QY 441 GAGAGAGAGCTTTTCCGCGCTGCGCCAGGGCTCTCCCGAGGGCGTGAACCGGGGGT 500  
DB 627 TCCGACAAAGATGAGAGGCGCGCTGCGCGCTTCTTCCGAGAGCGATGAGAGATGCT 686  
QY 501 GTGAGAACGAGCCCTCAATGAGTCTGGGAGCCAGCTGTGCTGCGGAAACGCGCCGTTG 560  
DB 687 CTGCGACCATG---TGATCTTCAAGCGCGCGCATCTGTCAAGCCGCAAGCGGGCTG 743  
QY 561 CGGGGCTGCGCCCTTAGGGCTTCTGCGGGGAGAGAGGCGCCCGGGCGCTACCCGCG 620

DB 744 CGGGGCTGCGCCCATCGCCCGCTCTGCGCGCTACGGGAGAGGCGAGACCGCA 803  
QY 621 GCCCAGAGAGCGCGGGCGAAGAGAGCGCTGTGCGCTGCTCC 667  
DB 804 GAGCGCAAGAGCTCTGAAGTACGAGAGGGCGGCTTCCCGGCG 850

RESULT 12  
US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 7.3%; Score 71.4; DB 14; Length 9025608;  
Best Local Similarity 47.1%; Pred. No. 2.6e-07;  
Matches 305; Conservative 0; Mismatches 321; Indels 21; Gaps 2;  
QY 39 GAAAAAGCCCGCCCTCCCTGCGGGGAGAGAGACCTTACCGCGTCTGCTC 98  
DB 5607423 GGTGTACCCGTACACACCCGAGTTGACTTGAGAACTCTTCCAGCTGATCTTGGC 5607482  
QY 99 CGAGGTCTTCTGAGAGAGACCGGGTGAAGCAGCCCTCCCTATTACCGCGCTTCT 158  
DB 5607483 GACCGTCTGTGCGCGAGACCACTGCGGTGAACCAAGACGCGCGCTGTT 5607542  
QY 159 GGAAGCGTTTCCACCTGAGAGCGCTGCGCGCTTCTTCCGAGAGAGTCTTAGAGT 218  
DB 5607543 CGCNAATACCGACCCCGAGAGACTGGCGCGCCCAACCCGAGAGGTGAGAGAT 5607602  
QY 219 CTGGCAGGGGGCGGCTACTACCGCGC-----GGCGAACACTTCCA 260  
DB 5607603 CTTCCGCGCCACCGGCTTCTCCGGGCCAAGACCAAGTCGTATAGGGTGTCCAAGGC 5607662  
QY 261 CCGCGTGGCCGAGAGCGTGAAGAGCTTCCCGAGCTTGGCGAGCTTCCGGGGCTTCC 320  
DB 5607663 CTTGTGAGAGATTGCGCGCGAGGTCCCGGGCGCTCGAAGATCTCTCAAGCTGCC 5607722  
QY 321 TGTCTCGGCTTACACCGCGCGCGGTGAGCTTCCATCGCTTCCGGGAGCGGGTGGC 380  
DB 5607723 CGGGTGAAGAGCAAGAGCGGCTTCTGTGTCTGGGCAACCTTCCGGCGCGCGCAT 5607782  
QY 381 GCGCGTGAAGAGAAAGTCCGAGAGGTCTTCTCCGCTCTTCCCGGGAAGCCCCAA 440  
DB 5607783 CACCGTGAACACCATTTTCCAGCGGCTGTCCGCGTGAAGTGAACCGAGAGAGA 5607842  
QY 441 GAGAGAGAGCTTTTCCGCGCTGCGCCAGGGCTCTCTCCCGAGGGCGTGAACCGGGGGT 500



```
Db      5607843 TC CGGACAAGATCGAGCGCGCGCTCTTCCCGAAGACGAGTGCATGCT 5607902
Qy      501 GTGGAACGAGCCCTTCATGAGCTCGGGCCACGGTCTTGCCTCGGAACGGCCCGCTTG 560
Db      5607903 CTCGCAACCTATG---TGATCTTCACGCGCGCGCGCATCTGTACACGCGCGGCGCTG 5607959
Qy      561 CGGGGCTGCGCCCTTAAGGGGCTTTTCCCGGGGAAGAGCCCGCGGCGCTTACCCCGC 620
Db      5607960 CGGGGCTGCGCCCTTAAGGGGCTTTTCCCGGGGAAGAGCCCGCGGCGCTTACCCCGC 5608019
Qy      621 GCCCAGGAAGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 667
Db      5608020 GAAGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5608066
```

## RESULT 13

```
US-10-398-221-1588
; Sequence 1588, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT FILING DATE: 2003-03-27
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1588
; LENGTH: 991
; TYPE: DNA
; ORGANISM: Listeria monocytogenes-4B
US-10-398-221-1588
```

Query Match 7.0%; Score 68.2; DB 15; Length 991;  
Best Local Similarity 49.7%; Pred. No. 7.9e-06;

Matches 215; Conservative 0; Mismatches 203; Indels 15; Gaps 1;

```
Qy      7 GCCTGGCGGAAGCCCTCTCGCTGCTGACCGGAAACGCCGCCCTCCCTCGGCGG 66
Db      77 GCTTTTCAGAGACACTGCTGTTCTCTGTATGAAGCGAATGAAGCGCTTTACCGTGGCGC 136
Qy      67 GGGGAGAAGAGACCTTACCGCGTCTGCTCCGAGTCTCTGAGAGAGAGAGAGAGAGAGAG 126
Db      137 GAATAATACGATCTTATCAGAAATTTGGGTCTCAGAAATCATGCTTTCACAAACAAAGTC 196
Qy      127 GAGCAGCCCTCCCTTATTAACCGCGCTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 186
Db      197 GACACAGTATTCCTTATTTATTCGTTATGACTCAATTTCCGACATGGAATATTTT 256
Qy      187 GCGCGGCTTCCCTGGAAGAGTCTTGAAGTCTGAGTCTGAGAGAGAGAGAGAGAGAGAG 246
Db      257 GTCCAGCGGAGAGAGAGAGAGATTTTAAAGCGTGAAGAGGTTAAGGCTTATTTCTGT 316
Qy      247 GCGGAACACCTC-----CACGCGCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 291
Db      317 GTCCAGAAACCTCCAAACAGCCATGAAGCAAGTGAAGGCGAATTTTTCGAGAAAGTTCT 376
Qy      292 CGAGGCTTCCCGAGAGTTCGGGGGCTTCTGCTCTCGGAGCTTACACCGCGCGGAGTGC 351
Db      377 ACCGACTTAACCACTATTTTATCGCTTAAAGGTGTGGAGCCGTATACAGCGGAGCTATC 436
Qy      352 GCGCTCATGCGCTTTCGGGAGAGCGGTGCGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAG 411
Db      437 TTGAGTATGCTTACCAACCAAGCTGAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 496
Qy      412 TCCCGGCTTCTTG 424
Db      497 GCCCGCGTTTAG 509
```

## RESULT 14

```
US-10-398-221-3596
; Sequence 3596, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT FILING DATE: 2003-03-27
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3596
; LENGTH: 3155
; TYPE: DNA
; ORGANISM: Listeria monocytogenes 4b
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3596
```

Query Match 7.0%; Score 68.2; DB 15; Length 3155;  
Best Local Similarity 49.7%; Pred. No. 6.2e-06;

Matches 215; Conservative 0; Mismatches 203; Indels 15; Gaps 1;

```
Qy      7 GCCTGGCGGAAGCCCTCTCGCTGCTGACCGGAAACGCCGCCCTCCCTCGGCGG 66
Db      713 GCTTTTCAGAGACACTGCTTCTCTGTATGAAGCGAATGAAGCGCTTTACCGTGGCGC 772
Qy      67 GGGGAGAAGAGACCTTACCGCGTCTGCTCCGAGTCTCTGAGAGAGAGAGAGAGAGAGAG 126
Db      773 GAATAATACGATCTTATCAGAAATTTGGGTCTCAGAAATCATGCTTTCACAAACAAAGTC 832
Qy      127 GAGCAGCCCTCCCTTATTAACCGCGCTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 186
Db      833 GACACAGTATTCCTTATTTATTCGCTTATGACTCAATTTCCGACATGGAATATTTT 892
Qy      187 GCGCGGCTTCCCTGGAAGAGTCTTGAAGTCTGAGTCTGAGAGAGAGAGAGAGAGAGAG 246
Db      893 GTCCAGCGGAGAGAGAGAGATTTTAAAGCGTGAAGAGGCTTGAAGTATTTCTGT 952
Qy      247 GCGGAACACCTC-----CACGCGCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 291
Db      953 GTCCAGAAACCTCCAAACAGCCATGAAGCAAGTGAAGGCGAATTTTTCGAGAAAGTTCT 1012
Qy      292 CGAGGCTTCCCGAGAGTTCGGGGGCTTCTGCTCTCGGAGCTTACACCGCGCGGAGTGC 351
Db      1013 ACCGACTTAACCACTATTTTATCGCTTAAAGGTGTGGAGCCGTATACAGCGGAGCTATC 1072
Qy      352 GCGCTCATGCGCTTTCGGGAGAGCGGTGCGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAG 411
Db      1073 TTGAGTATGCTTACCAACCAAGCTGAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1132
Qy      412 TCCCGGCTTCTTG 424
Db      1133 GCCCGCGTTTAG 1145
```

## RESULT 15

```
US-10-156-761-3629
; Sequence 3629, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
```



APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 3629  
LENGTH: 11058  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(11058)  
US-10-156-761-3629

Query March 6.9%; Score 66.8; DB 14; Length 11058;  
Best Local Similarity 48.9%; Pred. No. 9.6e-06;  
Matches 278; Conservative 0; Mismatches 272; Indels 18; Gaps 3;  
QY 331 CTTTACACCGCGCGCGCTGCTTCATCGCTTTGCGGAGACGGGTGCGCGGTGAC 390  
Db 6622 CCGCTCGCGTGCACGATCTACCGCGCTTCAGCAGAGCGGGCGCGCGCTCGAG 6681  
QY 391 GGGAACTCCGAGAGGTCTCTCCGCTCTTTCGCGGAAAGCCCAAGAGAGAG 450  
Db 6682 GAGTTCACTCGCCGCGGACCGCGCACCTGTTCACTGAGAGCGGCGCCGCTGCGG 6741  
QY 451 CTTTTCGCGCTCGCGCGCGCTCTCCCGAGGCGTGAACCGGGGGTGTGAACAG 510  
Db 6742 ATGACCGCACAGTGCAGAGAGAGAGCGCGCTCGCGCTTCAGCGATGTG--CCAC 6798  
QY 511 GCCCTCATGAGCTCGGGGCGCACGGTCTGCTGCGAAACGCGCCCGTTGCGGGGCTGC 570  
Db 6799 GCCATCACGAGAGGCGTGAAGCACCGCGCGCTGATGAGAGCTGCTGACGAAATACCG 6858  
QY 571 CCCCTAAGGAGCTTCTGCGGGGAGAGAGCGCTACCCGCGCGTACCGCGCCAGAG 630  
Db 6859 CGGCTGCGGACGCGGCGACAGCGCGTCCGAGAGCGCCCGCGCGCTACCGCGAG 6918  
QY 631 CGCGGCGAGAGAGAGCGCTGCTGCGCTCTGCTGCGAGCGGAGGGGGTGCAC 690  
Db 6919 TTCTGTGCGCACGAGGTGCTCTCTGAGTGCAGACGAGCGGCGTACTGCGGTGAC 6978  
QY 691 CTGAAAGGCTTGAAGGGGCGCTTCAG-----GGCTCTACGCGCGTCCCTCTTTC 742  
Db 6979 GTCAACCGCGCGCTTTCGCGCGCTTTCGCGCTGCGCGCGCGCGCGCGCGCTCGGC 7038  
QY 743 CCCCTGAGAGCTTCCCGGCGGAGAGCG-----CTTTCGAGTGAAGTCTAGGCC 795  
Db 7039 CCGCGGAGAGCTTCCGCGGTGCGCGTGAAGTGCAGCGCTTTCAGCAGAGGCGTGCAGCC 7098  
QY 796 CTAGCGAGGTGCGCGACCGCTACCGACGAGAGGCTTTCGCGTGAAGGTGCGGGGGCC 855  
Db 7099 CTGACCGCGCGCGCGCGGTCTCTCTCAAGAGCGTGTGCGCGCGCACCTGAAGCG 7158  
QY 856 CTTTGGAGAGGAGGAGGAGGAGCCCT 883  
Db 7159 CTGAGCGGTGACCGAGAGAGACCGCT 7186

Search completed: March 4, 2004, 07:02:34  
Job time : 1286 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 4, 2004, 04:02:01 ; Search time 76 Seconds  
(without alignments)  
7119.436 Million cell updates/sec

Title: US-09-938-901-1

Perfect score: 975  
Sequence: 1 gtgagagccgcgcgaagc.....tagtccccctccgcgaacca 975

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	125.4	12.9	963	4	US-09-252-991A-6832
2	125.4	12.9	1401	4	US-09-252-991A-6886
3	125.4	12.9	1413	4	US-09-252-991A-6739
4	113	11.6	1551	4	US-09-489-039A-4788
5	107.2	11.0	549	4	US-09-252-991A-6783
6	104	10.7	1858	4	US-08-813-574-1
7	87.2	8.9	2293	4	US-09-651-656-38
8	87.2	8.9	2293	4	US-09-650-855-38
9	86.6	8.9	4403765	3	US-09-103-840A-2
10	86.6	8.9	4411529	3	US-09-103-840A-1
11	82	8.4	9909	4	US-08-961-527-12
12	63.6	6.5	1212	4	US-09-107-532A-3654
13	63.6	6.4	1188	4	US-09-134-000C-1710
14	62.2	6.4	1227	3	US-09-385-028-23
15	62.2	6.4	1227	4	US-09-726-614-23
16	62.2	6.4	1227	4	US-09-385-040-23
17	62.2	6.4	15079	3	US-09-385-028-1
18	62.2	6.4	15079	4	US-09-726-614-1
19	62.2	6.4	15120	4	US-09-385-040-1
20	61.8	6.3	1044	4	US-09-543-681A-1143
21	61.4	6.3	1230035	4	US-09-198-452A-1
22	59.6	6.1	20235	1	US-07-642-734C-3
23	59.6	6.1	20235	1	US-08-439-009A-3
24	58.6	6.0	1041	4	US-09-328-352-856
25	58.6	5.8	2889	1	US-08-537-002A-4
26	56.6	5.8	2889	3	US-08-863-010-4
27	56.6	5.8	2889	3	US-09-024-429-4

28	56.6	5.8	3600	1	US-08-537-002A-5	Sequence 5, Appl1
29	56.6	5.8	3600	3	US-08-863-010-5	Sequence 5, Appl1
30	56.6	5.7	3600	3	US-09-024-429-5	Sequence 5, Appl1
31	55.6	5.7	2681	3	US-08-928-213B-7	Sequence 7, Appl1
32	55.6	5.7	2363	4	US-09-818-780-22	Sequence 22, Appl1
33	55.6	5.7	53500	4	US-09-266-965-76	Sequence 1, Appl1
34	55.2	5.7	2634	3	US-08-941-936-1	Sequence 1, Appl1
35	54.8	5.6	1404	4	US-09-252-991A-1725	Sequence 1725, Ap
36	54.8	5.6	2208	4	US-09-252-991A-2086	Sequence 2086, Ap
37	54.8	5.6	2586	4	US-09-252-991A-1894	Sequence 1894, Ap
38	53.8	5.6	3378	4	US-09-252-991A-2158	Sequence 2158, Ap
39	52.8	5.5	47981	4	US-09-679-279-1	Sequence 1, Appl1
40	52.8	5.4	4257	2	US-08-690-473-1	Sequence 1, Appl1
41	52.8	5.4	4257	3	US-09-821A-1	Sequence 1, Appl1
42	52.8	5.4	4257	3	US-08-843-659-1	Sequence 1, Appl1
43	52.8	5.4	12001	1	US-08-458-568A-11	Sequence 11, Appl1
44	52.6	5.4	2508	4	US-09-758-282B-48	Sequence 48, Appl1
45	52.6	5.4	3147	2	US-08-781-802-7	Sequence 7, Appl1

ALIGNMENTS

```
RESULT 1
US-09-252-991A-6832
Sequence 6832, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196, 136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 6832
LENGTH: 963
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6832

Query Match 12.9%; Score 125.4; DB 4; Length 963;
Best Local Similarity 54.7%; Pred. No. 2.1e-16;
Matches 311; Conservative 0; Mismatches 231; Indels 27; Gaps 2;

QY 55 CTCGCTGCGGCGGAGGAGGACCTTACCGGCTTCTGAGCGCTTCCACC 174
Db 53 CTCGCTGCGGCGGAGGAGGACCTTACCGGCTTCTGAGCGCTTCCACC 112
115 CAGACCGGCGGAGGAGGACCTTACCGGCTTCTGAGCGCTTCCACC 174
Db 113 CAGACCGGCGGAGGAGGACCTTACCGGCTTCTGAGCGCTTCCACC 172
175 CTGAGGCGGCGGCGGAGGAGGACCTTACCGGCTTCTGAGCGGCGGCG 234
Db 173 GTGAGGCGGCGGCGGAGGAGGACCTTACCGGCTTCTGAGCGGCGGCG 232
235 TACTACCGGCGGCGGAGGAGGACCTTACCGGCTTCTGAGCGGCGGCG 279
Db 233 TACTACCGGCGGCGGAGGAGGACCTTACCGGCTTCTGAGCGGCGGCG 292
280 GAGGAGCTTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 339
Db 293 GAGGAGCTTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 352
340 GCGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 399
Db 353 GCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 412
```

QY 400 CGAGAGGGTCTCTTCGCCGCTCTTTCGCCCGGGAGAAAGCCCCAAGAGAGAGAGCTTTTGGCC 459

Db 413 AAGCGCGCTACTGGCGCGCTACCTGGGCGCAGAGACGGCTATCCCGCGCAACCGAAGGTGGCC 472

QY 460 CTCGSCCAAGAGGCTCTCTCCCGCGAGGCGAGGACCCGGGGGTTGTGG-----AAC 507

Db 473 AGGGCGCTGTGGGAAGCCGCGCAAGCTTTCACCCCGACAGCAGCGGTTCAACCACTACACC 532

QY 508 CAGGCGCTTATGAGAGCTCGGGGCGACAGATCTGCTCTGCGAAGAACGGCCCGTTTGGCGGGCC 567

Db 533 CAGGCGATGATGGAACCTCGGCGCGCAACCTCTGCAAGCGCAAGAACCCAGTTGTCTGCTT 552

QY 568 TGGCCCCCTAGGGGCTTTTGTGCCGGGGGAA 596

Db 593 TGCCCGCTGGTCTCCGGCTGCGCGCGCA 621

```

RESULT 2
US-09-252-991A-6886/C
Sequence 6886, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ. ID NOS: 33142
SEQ. ID NO 6886
LENGTH: 1401
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6886

```

Query Match	12.9%	Score 125.4	DB 4	Length 1401
Best Local Similarity	54.7%	Pred. No. 2.2e-16		
Matches 311	Conservative 0	Mismatches 231	Indels 27	Gaps 2
QY	55	CTCCCTGCGGGGAGAGAGACCTTACCCGCTCTCGTCTCGAGGTCTCTTGAG	114	
Db	1137	CTGCCCTGGACAGAGGACATACCCCTACCGGGTGTGGTTCGGAAATCATCTGCAG	1078	
QY	115	CAGACCCGGGTGGAGCAGGCTCTCCCTATTACGCGCTTTCGAGCGCTTCCACC	174	
Db	1077	CAGACCCAGGTGAGCAGCCTGTGCTCGATTACTTCGACGTTTCATGAGCGGTGGCCAG	1018	
QY	175	CTGAAGCCCTGCGCGCGGCTTCCCTGGAAAGGCTCTTAAGGTCTGGCAGGGGCGGC	234	
Db	1017	GTGAGGACATGCGCCGCGCGCGCGAGACGAAAGTCTGCACTGTGGACGGGCTGGC	958	
QY	235	TACTACCGGCGGGCGGAAACCTCTCACCGCTGCGCCGAGACGT-----G	279	
Db	957	TACTACACCGCGGCGCGCACTTGCAAGAACCGGACGATCTGTGTGTCAGCGGCATGCG	898	
QY	280	GAGAGGCTTCCCCCGAGCTTGCAGAGTTGGGGGCTTCTGTCTTGGGCTTACACC	339	
Db	897	GGGAGTTCCTCCCCCGAGCTGAGACATCTGCCAATCGCCCGGCATCGCGCTCCACC	838	
QY	340	GCGGCGGCGGTGCGCTTCATCGCTTCGGGGAGGGGGGGGGGGGGGAGCGGAAGTC	399	
Db	837	GCTGAGGCAATCGGCACGCTGTGCATGAGCTTGCGCGCAGCAGATCTTCGACGGGAAGTC	778	
QY	400	CGAGGGTCTCTCCCGCTTTCGCCCCGGGAAAGCCCCAAGAGGAAGAGCTTTTCGC	459	
Db	777	AAGGCGGACTGCGCGCTACTGTGGCGGAGGAGCGGCTATCCCGCGAAGCGAAGTGGCC	718	
QY	460	CTGGCCAGGCGCTCTCTCCCGAGGGGTGAGATCCGGGGGCTTGG-----AAC	507	

Db 71, AGGGCGCTGTGGGAAAGCCCGGACAGCTTCAACCCCGCAGCAGCGGTCAACCACTACACC 658

Qy 508 CAGGCGCTCATGAGAGTGTGGGGGCCACGSGTCTGCTGTCCGAAACGGCCCCGTTCGGGGCC 567

Db 657 CAGGCATATATGAACTCTGGGCGCCACCTCTTGCAACGGCAGCAAGACCCAGTTGCTCTT 598

Qy 568 TGCCCCCTTAGGGGCTTGTGTCGGGGGAA 596

Db 597 TGCCGCTGTCTCCGAGCTGCGCGCGCA 569

```

RESULT 3
US-09-252-991A-6739
; Sequence 6739, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074.788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094.190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6739
; LENGTH: 1413
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-6739

```

Query Match	12.9%	Score 125.4	DB 4	Length 1413
Best Local Similarity	54.7%	Pred. No. 2.2e-16		
Matches 311	Conservative 0	Mismatch 231	Indels 27	Gaps 2
QY	55	CTCCCTTGCGCGGGGAGAAAGAACCTTATCCGCGGCTCGTCCGAGGTCTCTTCAG	114	
Db	409	CTGCCTTGCGACGAGGGGATCACTCCCTTACCGGCTGGGCTCTGGAAATTATCTCGAG	468	
QY	115	CAGACCCGGGTGGAGACAGGCGCTTCCCTATTATCCGCGCTTTCGAGCGCTTTCGAC	174	
Db	469	CAGACCCGAGTTCAGACCGGTGCTGGTTACTTTCGACCGTTTCATGGACGGGTGCCGAC	528	
QY	175	CTGAAGCGCCCTGGCGCGGGCTTCCCTGGAAGAGTCTTATAGGCTCTGGACGGGGCGGC	234	
Db	529	GTCAAGGACTCTGGCGGGCGGCGGAGAGAAATCTTCGACCTGTGTGACCGGGCTCGGC	588	
QY	235	TACTACCGCGCGGGGAGAAACCTTCACGCGCTGGCCCGAAGCGT-----G	279	
Db	589	TACTACGACCGGTGGGCCCACTGTCAAGAACCGCGAGATCGTGTGCGACGGCATGCG	648	
QY	280	GAGAGCTTCCCCCGAGCTTTCGCGAGCTTCGAGGCGCTTCTGTGCTCGGCGCTTACAC	339	
Db	649	GGGAGTTCCCCCGCGCAGTTCGAGCAACTCGCGGAATCGCCGCAATCGGCCCTCCACC	708	
QY	340	GGCGCGCGGGTGGCTTCATCGCTTCGCGGAGCGGTGGCGCGGTGAGCGGAACGTC	399	
Db	709	GCTGAGAACCAATCCGACAGCTGTGATAGGGCTTCGGGCAACCGATCTTCGACGCGCAACGTC	768	
QY	400	CGAGAGGTCTCTCCGCGCTTCGCGCGGGAAAGCCCAAGGAGAAAGAGCTTTTCGCC	459	
Db	769	AAGGCGGTATAGCGCGGCTTACTGTGGCGCAGAGAGCGCTATTCGCGCAACCGAAGTGGCC	828	
QY	460	CTGCGCCAGGCGCTCTCTCCCGAAGGCGTGAACCCGCGGGGTGG-----AAC	507	
Db	829	AGGGCGCTGTGGGAACCGCGCAACGCTTCACCCCGCAGCAACGGGTCAACCACTACAC	888	
QY	508	CAGGCGCTATGAGAGCTCGGGGCTCAAGGTCTGCTTCGCAAAAGCGCCCGTTGGCGGGCC	567	
Db	889	CAGGCATATATGACCTTGCGCGCACTCTCTGACCGCGCAGAACGACGATGGCTGCTT	948	

Qy 568 TGCCCTGAGGGGCTTGTGCGGGGGA 596  
 Db 949 TGCCCTGAGTCTCGGGCTCGGGCGCA 977

RESULT 4

US-09-489-039A-4788  
 ; Sequence 4788, Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Brecon et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 2709,2004001  
 ; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; PRIOR FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 4788  
 ; LENGTH: 1551  
 ; TYPE: DNA  
 ; ORGANISM: Klebsiella pneumoniae  
 ; US-09-489-039A-4788

Query Match 11.6%; Score 113; DB 4; Length 1551;

Best Local Similarity 53.0%; Pred. No. 6,1e-14;

Matches 307; Conservative 0; Mismatches 245; Indels 27; Gaps 2;

Qy 53 CCCCTCCCTGCGGGGAGAGAGACCTTACCGGTCCTGCTCGAGGCTCTTGC 112  
 Db 560 CCTTCCCTGCAATGATGCAAGAGCCCTTCAAGATGAGCTCTCCAGATGATGTC 619  
 Qy 113 AGCAGACCCGGGTGAGAGAGCCCTTCCCTTATTAACCGCGCTTCTGAGGCTTTCCA 172  
 Db 620 AAGAAACCAAGTACACACCGTATCCCTTATTTGAAGCTTTATGCGCGTTTCCCA 679  
 Qy 173 CCTGAAAGCCCTGCGCGGCTTCTGAGAGGCTCTTAAAGGCTGCGAGGGGCGG 232  
 Db 680 CGGTGATGATCTGCGCAACCGCGCTGATGATGATGATGATGATGATGATGATG 739  
 Qy 233 GCTACTACCGGCGGGGGAACAC-----CTCACCGCTGCGCGGAAGCG 277  
 Db 740 GCTACTACCGGCGGGGCGGCACTTACATTAAGCCGCGGACAGACGCTCCACATGCA 799  
 Qy 278 TGAAGAGCTTCCCGAGCTTCCCGAGCTTCCGGGCTTCTGATCTCGGCTTACA 337  
 Db 800 GCGGGGAATCCCGGAGCTTTGACGAAGTGGGGGCTACCGGGGCTGGGGGCTCA 859  
 Qy 338 CCGCGGCGGCGGCTTCCATCGCTTTCGGGAGAGCGGCGGCGGCTGAGCGGAACG 397  
 Db 860 CCGGCGGCGGCTTATTCCTTTCGCTGCTGAGCATTCATTCGATTCGACCGCAACG 919  
 Qy 398 TCCGGAAGGTCCTTCCCGCTTTCGCGCGGAAAGCCCAAGAGAGAGCTTTTCG 457  
 Db 920 TGAAGCGGCTGCTCGCGCTGATGTCAGCGGCTGCGCGGAAAGAGGTCG 979  
 Qy 458 CCTCGCCAGAGGCTCTTCCCGAGGCGTGAACCGGCGGCTTTCG-----A 505  
 Db 980 AAAAAGCGCTGTGGGATACAGCAAGAGTCAACCGCGGAGAGGCGCTGAGCGCTTCA 1039  
 Qy 506 ACCAGGCTTCATGAGCTGCGGCGCAAGGCTGCTGCGGAACGCGCGCTTTCGCGGG 565  
 Db 1040 ACCAGGCGATGATGATCTCGGGGCAATGATTTGACCCGCTCGAAGCGAAGTGCAGC 1099  
 Qy 566 CCGGCGGCTGAGGCGCTTTCGCGCGGGAAGAGGCGCC 604  
 Db 1100 TGTGTCCGTGAGCAAGCGCTGCTGCTTACGGAAC 1138

RESULT 5  
 US-09-252-991A-6783  
 ; Sequence 6783, Application US/09252991A

Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubinfeld et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196,136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 6783  
 ; LENGTH: 549  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-6783

Query Match 11.0%; Score 107.2; DB 4; Length 549;

Best Local Similarity 54.4%; Pred. No. 7.6e-13;

Matches 248; Conservative 0; Mismatches 193; Indels 15; Gaps 1;

Qy 55 CTCCCTGCGGGGAGAGAGACCTTACCGGCTCTGCTCGAGGCTCTTTCGAG 114  
 Db 60 CTGCTCTGCGAGCGAGGCTATCCCTTACCGGCTGAGGCTCTGGAATATGCTGAG 119  
 Qy 115 CAGACCGGCGTGAAGAGCCCTTCCCTTATTAACCGCGCTTCTGAGCGCTTTCCACC 174  
 Db 120 CAGACCGGCGTGAAGAGCCCTTCCCTTATTAACCGCGCTTCTGAGCGCTTTCCACC 179  
 Qy 175 CTGAAGGCTCTGCGGCTTCTTCTGAGAGGCTCTTGAAGGCTCTGAGGCGGCGG 234  
 Db 180 GTCGAGGCTCTGCGGCTTCTTCTGAGAGGCTCTTGAAGGCTCTGAGGCGGCGG 239  
 Qy 235 TACTACCGGCGGAGAGACCTTACCGGCTTCCCGGAGAGCTT-----G 279  
 Db 240 TACTACCGGCGGAGAGACCTTACCGGCTTCCCGGAGAGCTT-----G 279  
 Qy 280 GAGGAGCTTCCCGGAGCTTTCGAGGCTTTCGGGCTTCTGAGGCTTTTACAC 339  
 Db 300 GGGAGTTCCTCCCGGAGCTTTCGAGGCTTTCGGGCTTCTGAGGCTTTTACAC 359  
 Qy 340 GCGGCGGCGGCTTCTTCTGAGGCTTTCGGGAGGCGGCTTTCGGGAGGAGGCTC 399  
 Db 360 GCTGAGGCTTTCGAGGCTTTCGGGAGGCGGCTTTCGGGAGGAGGAGGCTC 419  
 Qy 400 CGGAGGCTTTCGAGGCTTTCGGGAGGCGGCTTTCGGGAGGAGGAGGAGGCTTTCGCC 459  
 Db 420 AAGCGGCTTTCGAGGCTTTCGGGAGGCGGCTTTCGGGAGGAGGAGGAGGCTTTCGCC 479  
 Qy 460 CTCGCCAGGCGCTTTCGCGGAGGCGGCTTTCGCGGAGGCGGCTTTCGCGG 495  
 Db 480 AGGCGGCTTTCGCGGAGGCGGCTTTCGCGGAGGCGGCTTTCGCGG 515

RESULT 6

US-08-813-574-1  
 ; Sequence 1, Application US/08813574

Patent No. 6013473

GENERAL INFORMATION:

APPLICANT: Wei, Ying-Fei

TITLE OF INVENTION: Human Mucy

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

```
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS Windows Version 2.0
SOFTWARE: FastSeq for Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/813,574
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/013,132
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: ATG50002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1858 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-813-574-1
```

Query Match 10.7%; Score 104; DB 3; Length 1858;

Best Local Similarity 54.0%; Pred. No. 3.6e-12;

Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;

```
QY 69 GGAAGAAGACCTTACCGGCTCTGCTCCGAGCTCTTCTGACAGACCGGAGTGA 128
DB 498 GGAAGACGCGGCTATGCTGTGTGCTGAGGCTGCTGACACCAACCGAGTTGC 557
QY 129 GCAGGCGCTCCCTATTTACCGCGCTTTCTGAGCGCTTTCCACCTGAAGCGCTGAG 188
DB 558 CACTGATCACTACTATACCGGATGATGCAAGATGAGCTCACTGACAGACCTGAGC 617
QY 189 CGGCGCTTCTGGAAGAGTCTTGAAGGTCTGCGAGGCGGCGGCTACTACCGCGAGC 248
DB 618 CAGTGTCTCTGGAAGAGTGAATCACTGTGGCTGAGCTGAGCTATTTCTGTGG 677
QY 249 GGAACCTTCCACCGGCTGCGGAGCGTGAAGAGCTTCCCGAGCTT----- 300
DB 678 CCGGCGGCTGCAAGAGAGCTGGAAGGTGTAAGAGCTAAGGAGCCACATGCGACG 737
QY 301 -----GCCGAGCTTCGAGGCGCTTCTGAGCTCGGAGCTTACACCGCGCGAGT 350
DB 738 TACAGAGAGACCTTGCACAGCTCTGCTGCGGCGCTGACACAGCTGGGCTCAT 797
QY 351 GGCCTTCATCGCTTCTGAGAGCGGAGCGGCGGCTGACCGGAACGTCCGAGAGGTCT 410
DB 798 TGCCTTATCGCTTGTGGCCAGGCAACCGATGTGTGATGCAACGTAGCACGGGTCT 857
QY 411 CTCGCGCTCTTCGCGCC--GGGAAGCCCAAGAGAGAGAGCTTTCCGCTCGCCGAG 468
DB 858 GTGCGGTGTCGAGCATTTGTGTGCTGATCCAGACACCTTGTTCACAGAGTCTG 917
QY 469 GGCCTTCCTCCGAGGCGCTGACCG-----GGGATGTGAAACAGGCGCTCAT 518
DB 918 GGGGTAGCCACGACAGCTGTGGAACCCAGCCGCGCAGAGATTCAACAGAGAGCT 977
QY 519 GGAAGCTCGGAGCGACGCTGTGCTGCGGAACGAGCGCGCTGCGGAGCTGCGCCCTAG 578
DB 978 GGAAGCTAGGAGCGACAGTGTGTAACCCACAGGCGCACAGTGTGAGCTGCTGTGA 1037
QY 579 GGCCTTCTGCGGAG 592
DB 1038 GAGCTGTGCGGAG 1051
```

RESULT 7

```
US-09-651-656-38
Sequence 38, Application US/09651656
Patent No. 6340566
GENERAL INFORMATION:
APPLICANT: MCCUTHEN-MALONEY, SANDRA
TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE
POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,
FILE REFERENCE: IL-10689
CURRENT APPLICATION NUMBER: US/09/651,656
PRIOR FILING DATE: 2000-08-29
PRIOR APPLICATION NUMBER: 60/192,764
NUMBER OF SEQ ID NOS: 106
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38
LENGTH: 2293
TYPE: DNA
ORGANISM: Escherichia coli
US-09-651-656-38
```

Query Match 8.9%; Score 87.2; DB 4; Length 2293;

Best Local Similarity 54.1%; Pred. No. 7.2e-09;

Matches 210; Conservative 0; Mismatches 163; Indels 15; Gaps 1;

```
QY 53 CCTTCCTTGGCGGGGAGAGACCTTAACCGGCTCTGCTCGAGGCTCTTCTGC 112
DB 1161 CTCTCCCTCGCAATTGACAAAGCCCTCAAAAGTATGCTCAGAAATGATGTTGC 1220
QY 113 AGCAACCGGGGTGAGAGAGCCCTTCTTACCGCGGCTTCTGAGGCGTTTCCCA 172
DB 1221 AACCAACTGAGTGCAGCCTTATCCCTTATTTGAAAGCTTATGAGCGGCTTCCGA 1280
QY 173 CCTGAAGCGCTGCGCGGCTTCTCTGAGAGAGTCTTATAGGTCTGCGAGGAGCGG 232
DB 1281 CGGTACCGATCTGCGCAATGCGCGCTGACAGAAAGTCTCAGTGTGAGACCGGCTTG 1340
QY 233 GCTACTACGGGCGGAGAACCTTCAACCGCTGCGCGAGCGTGAG----- 282
DB 1341 GCTATTACGCGCGCGCGCAATCTGCAATAAGCGGCAACAAGTGGCGACCTTACAG 1400
QY 283 -----GAGCTTCCCGGAGCTTCCGCGAGCTTCCGAGGCTTCTGAGGCTTACA 337
DB 1401 GCGGTAATTCGGAACCTTTGAGAAATTGACACTCGCGGCTCGGAGCTTCCA 1460
QY 338 CCGCGCGCGAGTGTGCTTCATCGCTTCTGAGAGCGGAGTGCAGCGGTGAGCGGAAG 397
DB 1461 CCGCAGGCGGATTTCTGCTTCTGAGTGAAGCACTTCCGATTTCTGACGAGTAACG 1520
QY 398 TCCGAGGAGTCTTCTCCGCTCTTTCG 425
DB 1521 TCAACGCGGTCTGCGCGCTGCTATGC 1548
```

RESULT 8

US-09-650-855-38

Sequence 38, Application US/09650855

Patent No. 6365355

GENERAL INFORMATION:

APPLICANT: MCCUTHEN-MALONEY, SANDRA

TITLE OF INVENTION: LIVERMORE NATIONAL LABORATORY

TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA

TITLE OF INVENTION: MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA

FILE REFERENCE: IL-10284

CURRENT APPLICATION NUMBER: US/09/650,855

PRIOR FILING DATE: 2000-08-29

PRIOR APPLICATION NUMBER: 60/192,764

NUMBER OF SEQ ID NOS: 106

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 38

LENGTH: 2293  
 TYPE: DNA  
 ORGANISM: Escherichia coli  
 US-09-650-855-38

Query Match  
 Best Local Similarity 54.1%; Pred. No. 7.2e-09;  
 Matches 210; Conservative 0; Mismatches 163; Indels 15; Gaps 1;

53 CCTCTCCCTGGCGGGGAGAGACCTTACCGGCTCTGCTCCGAGGCTCTTCTGC 112  
 1161 CTCTGCTGCAATTTGACAGACGCTCTACAAAGTAGCTCTCAGAAATGATGTTGC 1220  
 113 AGCAGACCCGGGTGAGAGAGCCCTCTCTTATACCGGCTTCTGAGAGCTTTCCCA 172  
 1221 AACAACTCAGGTGGCCGACCTTTATCCCTTATTTGAAGCTTTATAGCGGCTTCCGA 1280  
 173 CCTGAAGAGCCCTGCGCGGCTTCCCTGGAAGAGTCTTAGGGTCTGCGAGGGGCGG 232  
 1281 CGGTGACCGATCTGCGCAATGCGCGCTCGACGAATTTCTCACTTGTGACCGGCTTG 1340  
 233 GCTACTACCGCGCGGCGGACCTTCAACCGCTGCGCGGCGGAGCGTGGAG----- 282  
 1341 GCTATTACCGCGCGGCGGCAATCTGATTAAGCGGCAACAGTGGCGACTTACACG 1400  
 283 -----GAGCTTCCCCAGCTTCCCGGAGCTTGGGGGCTTCTGCTGCTTACG 337  
 1401 GCGGTAAATTCCTCGGAACTTTGAGAGATTGACAGCTCCGCGGCTCGGGGCTTCCA 1460  
 338 CCGCGGCGCGGCGGCTTCCATCGCTTCCGAGAGCGGGGTGCGCGGCTGAGCGGAAAG 397  
 1461 CCGGAGGCGGATCTCTGCTTCTCTGCGTAAACACTTTCGATTTCCAGCGGTACG 1520  
 398 TCCGAGGCTCTCTCCGCTCTTCCG 425  
 1521 TCAGACGCTGCGCGGCTGCTATGC 1548

RESULT 9  
 US-09-103-840A-2  
 Sequence 2, Application US/09103840A  
 Patent No. 6294328  
 GENERAL INFORMATION:  
 APPLICANT: FLEISCHMAN, Robert D.  
 APPLICANT: WHITE, Owen R.  
 APPLICANT: FRASER, Claire M.  
 APPLICANT: VENTER, John C.  
 TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 FILE REFERENCE: 24366-2007.00  
 CURRENT APPLICATION NUMBER: US/09/103.840A  
 CURRENT FILING DATE: 1998-06-24  
 NUMBER OF SEQ ID NOS: 2  
 SOFTWARE: Patent Ver. 2.1  
 SEQ ID NO 2  
 LENGTH: 4403765  
 TYPE: DNA  
 ORGANISM: Mycobacterium tuberculosis  
 FEATURE:  
 OTHER INFORMATION: CDC 1551  
 OTHER INFORMATION: "n" bases at various positions throughout the sequence  
 OTHER INFORMATION: represent a, t, c or g  
 US-09-103-840A-2

Query Match  
 Best Local Similarity 8.9%; Score 86.6; DB 3; Length 4403765;  
 Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;  
 22 CTCCTGCGCTGTACCGGAAAGCGCCCTCTCCCTGCGCGGG--GAGAGAGAC 78  
 4022631 CTCTGCGCTGTATCAGCGATCCGACCGGATCTGCGCTGCGAGACCGGCTGTCAGC 4022690  
 79 CTTTACCGGCTGTCTCGAGGTCTTCTGACAGACCGCGGTGAGACGCGCTC 138

Db 4022691 CCGTGCAGATCTGTCTAGGATTCATGTGACAGAGACCGCGCGCGGTGCTG 4022750  
 139 CCCTATTACCGCGCTTCTGAGACGCTTCCACCTGAAGCCCTGCGCGGCTTCC 198  
 4022751 GCGATTCGCGCGGATCTGCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 4022810  
 199 CTGAAGAGGCTCTTAGGTCTGCGAGCGGCGGCGCTACTACCGCGCGGCGGAAACACTC 258  
 4022811 ACCGCGATGTGTAACGCGCTGCGGCGAACTGCGCTATCCAGCGCGCAAGCGCTTA 4022870  
 259 CACCGCTGCGCGCGAGCGT-----GAGAGCTTCCCGGAGCTTCCG 303  
 4022871 CACGATGCGCGCACGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4022930  
 304 GAGCTTGGGGGCTTCTGCTGCTGCGGCTTACACCGCGCGCGCGGTGCTCATCGCC 363  
 4022931 ATCTGTGACACCTGCGCGGCGTGGGAGCTACACCGCGCGCGCGGTGCTGTTGCT 4022990  
 364 TTGCGGAGCGGCTGCGCGGCTGAGCGGAACTCCGGAAGGTCTCTCCGCTCTT- 422  
 4022991 TACGCGAGCGGCTGCGGCTGAGACCAATGTGCGCGCGGTGCGCGCGCTT 4023050  
 423 -----CGCGCGGAAAGCCCAAGAGAGCTTTTCCGCTCGCGCGAGGAC 471  
 4023051 CACG 4023110  
 472 CTCCTCCCGAGGCGGTGACCGCGGCGGTGGAACAGCGCTCATAGAGCTCGGGCC 531  
 4023111 CTGTGCGCGGACCGCGAGACCGCGCTGAAATTTTCGCTGCGGTGATGAGTGGGTGCG 4023170  
 532 ACGGTCTGCTGCGGAAAGCGCGCTTGGCGGCGCTGCGCTTACGAGGCTTCTGCGCG 591  
 4023171 ACGGTGTGACACCGCGCGCGCGCGCGCGGTGCGGTATGCTCGGTGAGCTGTGCGCATG 4023230  
 592 GGAAGAGGCGCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 651  
 4023231 CCGCATGCGGTATTCGCGCGCTCGGAGCGTCC---GCCGCGCGCGGCGAGCGCTACACC 4023287  
 652 CTCGTGCGCTGCTGCTCTCTGCGGCGG 678  
 4023288 GGAACCGACCGCGCAATCTCGCGGACGG 4023314

RESULT 10  
 US-09-103-840A-1  
 Sequence 1, Application US/09103840A  
 Patent No. 6294328  
 GENERAL INFORMATION:  
 APPLICANT: FLEISCHMAN, Robert D.  
 APPLICANT: WHITE, Owen R.  
 APPLICANT: FRASER, Claire M.  
 APPLICANT: VENTER, John C.  
 TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 FILE REFERENCE: 24366-2007.00  
 CURRENT APPLICATION NUMBER: US/09/103.840A  
 CURRENT FILING DATE: 1998-06-24  
 NUMBER OF SEQ ID NOS: 2  
 SOFTWARE: Patent Ver. 2.1  
 SEQ ID NO 1  
 LENGTH: 4411529  
 TYPE: DNA  
 ORGANISM: Mycobacterium tuberculosis  
 OTHER INFORMATION: H37Rv  
 US-09-103-840A-1

Query Match  
 Best Local Similarity 8.9%; Score 86.6; DB 3; Length 4411529;  
 Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;  
 22 CTCCTGCGCTGTACCGGAAAGCGCCCTCTCCCTGCGCGGG--GAGAGAGAC 78

Db 4030550 CTTCTGCTGGTATCAGGATCCGACCGGAGATCTGCCCCGACGAGACCCGGGTCTCAGC 4030609  
Qy 79 CCTTACCCGCTCTCTGCTCCGAGGTCTTCTGACGACGACCCGGGTGGAGCAGCCCTC 138  
Db 4030610 CCGTGGCAGATCTCTGGTCAAGGATTCATGCTGACGACGACGCGCCGGGTCTG 4030669  
Qy 139 CCTTATTAACCCGCTTCTGAGAGGCTTTCCACCCCTGAAGGGCCCTGGCCGGCTTC 198  
Db 4030670 GCGATCTGGCCGAGTGGGTGGCGGTGGCCGACCCCTGGCCGACCGCCAGCAGC 4030729  
Qy 199 CTGGAAGAGGTCTTAAAGGTCTGACGAGGGGCGGCTTACTACCGCGGCGGAAACACTC 258  
Db 4030730 ACCGCGATGTGTAAAGCGCTGGGGCAAGCTGGGCTATCCAGCGGACCAAGCGTTA 4030789  
Qy 259 CACGCGCTGGCCGAAAGCTG-----GAGAGCTTCCCGCAGCTTGGCC 303  
Db 4030790 CACGATGCGGCACCGGTATGCGCCGACCAATATACGTGGTGGCCGACGATATGAG 4030849  
Qy 304 GAGCTTGGGGGCTTCTGCTGCTGCGGCTTACACCGCGCGCGGTGGCTCCATGCGC 363  
Db 4030850 ATCTGTGTCACCTGGCGGGGCTGGGAGCTACACCCGCGCGGTGGCTTTCCT 4030909  
Qy 364 TTCCGGAGCGGGTGGCGCGGTGAGAGGAACTCCGAGAGGTCTTCTCCGCTCTT 422  
Db 4030910 TACCGCAGCGGGTGGCGGTGAGACCAATGTGCGCGGTGGTGGCCGCGCTT 4030969  
Qy 423 -----CGCCGCGGAAAGCCCAAGAGAGAGCTTTTCCGCTCGCCGAGGCG 471  
Db 4030970 CACGCGCGCGCCGACCGCGGTGGCGCTATCGGTCCGCGCAACACGCGCTTGGCG 4031029  
Qy 472 CTCTCTCCCGAGGCGGTGACCGCGGGGTGTGAACAGGCTCATGAGCTCGGCGC 531  
Db 4031030 CTGTTGCGGACCGCGAGAGCGGCTGAATTTTCGCTGCTGATGAGTTGGGTGG 4031089  
Qy 532 ACCGCTGCTGCGGAAAGCGCCCTTGGGGGCGCTGCCCCCTTAAAGGCTTCTCGCG 591  
Db 4031090 ACCGTGTGACCGCCGCGCAACCCCGGTGGGTATGCTCGCTGAGCTGGTCCGATG 4031149  
Qy 592 GGGAGAGAGGCCCCCGGCGCTACCCCGCGCCGACGAGAGCGCGGAGAGAGAGCGG 651  
Db 4031150 CGGCATGCGCGTTATCGCGCGTGGAGCGTCTC---GCCGCGCGGGGGGAGGCTTACAC 4031206  
Qy 652 CTGCTGCGCTCTGCTCTCTCGGCGCG 678  
Db 4031207 GGAACGACCGCCCAAGTCCGCGAGCG 4031233

## RESULT 11

US-08-961-527-12/C  
Sequence 12, Application US/08961527  
Patent No. 6420135  
GENERAL INFORMATION:  
APPLICANT: Charles Kunesh  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9909 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-12

Query Match 8.4%; Score 82; DB 4; Length 9909;  
Best Local Similarity 52.2%; Pred. No. 8.6e-08;  
Matches 218; Conservative 0; Mismatches 185; Indels 15; Gaps 1;

Qy 22 CTCTCGCTGTACCGGAAACGCCCCCTCCCTGGCGGGGAGAGACCT 81  
Db 5808 CTCTCGCTGTATGATGAAACAAAGAGATTTCCTTGAGAGAGAAATCT 5749  
Qy 82 TACCGGCTCTGTCTCCGAGTCTTTCGACGACCGGGTGGAGAGCCCTCCC 141  
Db 5748 TATCATCTGGATATCTGAATCATGCTTACGACACGAGGTGATACATATCT 5689  
Qy 142 TATTACCCGCTTCTGAGAGCTTTCACCTGAGAGCCCTGAGCGCGGCTTCCCTG 201  
Db 5688 TACTACGAAAGATTTTGTGACTGTTTCACTGTGCAAACTGCGCACTGCGCTGAG 5629  
Qy 202 GAAAGAGCTTGAAGGTCTGCGAGGGGCGGCTA-----CTACCGCGG 246  
Db 5628 GAGAGTTACTGAAAGCTTGGAGGGGCTTGGCTATTATTCTCGAGTTCCCAATATGAG 5569  
Qy 247 GCGGAACTCTCAACCGCTGCGCGCAAGCGTGAAGAGCTTCCCGAGCTTCCGAG 306  
Db 5568 GCTGACGCCGACGATTTAGTACTGACTTGTGGCCAAATTTCCAAATCTATGAGAG 5509  
Qy 307 CTGCGGGGCTTCTGTCTCGGCTTACACCGCGCGCGGTGCGCTTCATGCGCTTC 366  
Db 5508 ATTTCAGCTTGAAGAGGATTTGACCTTACACGACGAGACCAATTCAGATTTGCTTT 5449  
Qy 367 GGGAGAGGGGTGGGGGTGGAGAGGAGAGTCCGAGAGGTCTCTCCGCTTTG 424  
Db 5448 AACTTGCTGAGCCAGCTGTGATGTATGTATGCGGGTCTTGGCGGCTGTTTG 5391

## RESULT 12

US-09-107-532A-3654  
Sequence 3654, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998



## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arieliello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 3654:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1212 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (8) LOCATION 1...1212  
SEQUENCE DESCRIPTION: SEQ ID NO: 3654:  
US-09-107-532A-3654

Query Match 6.5%; Score 63.6; DB 4; Length 1212;

Best Local Similarity 47.6%; Pred. No. 0.00029;  
Matches 278; Conservative 0; Mismatches 279; Indels 27; Gaps 2;

31 TGGTACCGGGAAGCGCCGCCCTCCCTGCGGGGGGAGAGACCTTACCGCGCTC  
100 TGGATGACAGAAAGAAAGAAACCTGCTGGGCGTATATCGATCGATTACGATC  
91 CTGCTCTCCGAGGCTCTTCTGACAGACCGCGGTGAGCAGCGCCCTTATACCGC  
160 TGGATCTCTGAGATCATGCTTCAGCAACAGAGTAGACAGATGATGATTTAT  
151 CGCTTCTGAGCGCTTCCACCGTGAAGCGCGCGGCTCCCTGGAAGAGTGC  
220 CGTTTCATGAGATGATTTCTCAATGAAAGCTAGCAAAATGCTCCGAAAGAAAGCTT  
211 CTTAGGCTCTGCGACGCGCGGCTACTACCGCGCGGCAACACTC-----  
280 TTAAGAGATGAGGAGCGCTTGTATTACTAGAGCGCGGAACATCAAGCTGCT  
259 --CAACGCTGCGCCGAGCGTGAAGAGCTTCCCGAGCTTCCGAGCTTGGGGG  
340 AATACAGATCATGATGAAATTTGATGGGGAATGCGCAACGCTGAAGAAATTAGTCA  
316 CTTCCTGCTCTGCGGCTTACACCGCGCGCGGTGCTTCATCGCTTGGGGAGCG  
400 TTGAAGGAATAGAGCATATATACAGACAGCGCGATGAGATCGCATTTGGAATTCCA  
376 GTGCGCGCGTGAAGCGGAACGTCCGAGAGGCTCTTCCGCTCTTCCGCGGAAAGC  
460 GAACCTGCAAGTTGACGCGACGATGAGAGTATGAGATGTTTGGATTTGAGCA  
436 CCCAAGAGAA-----GAGCTTTTGGCCCTCGCCAGGCGCTTCCCGAG 483  
520 GACATCCCAAAAGCTTCTAGCAGAGAAATCTTTGATGAAAGCATGCGGAATATGAT 579  
484 GGCCTGACCCGCGGGGTGGAACCAAGCCCTCATGAGCTCGGGGCAAGCTTGGCTG 543  
580 GAAAGACATCAAGGTGATTAACCAACAGCATGATGATCTAGGTTACGATCTGTACA 639  
544 CCGAAGCGCGCGCTTGGCGGGCGCTCCCTAGGCGCTTCTG 587  
640 CCAACTCTCCAAAGTGAAGATTTGTCGATCCAGCTTTTG 683

## RESULT 13

US-09-134-000C-1710  
Sequence 1710, Application US/09134000C  
Patent No. 6617156  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucelte-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 032796-032  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US/09/134,000C  
PRIOR FILING DATE: 1997-08-15  
NUMBER OF SEQ ID NOS: 6812  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1710  
LENGTH: 1188  
TYPE: DNA  
ORGANISM: Enterococcus faecalis  
US-09-134-000C-1710

Query Match 6.4%; Score 62.4; DB 4; Length 1188;

Best Local Similarity 48.8%; Pred. No. 0.0005;  
Matches 211; Conservative 0; Mismatches 206; Indels 15; Gaps 1;

8 CTTGGCGGAAAGCCCTCTGCTGTACCGGGAAGCGCCGCCCTCCCTGGCGGG 67  
53 CATTCCAGAAAGATTTTATGCTGTATGACGAGAAAGCAATTTACCTTGGCGAG 112  
68 GGGAGAGAGACCTTACCGCGCTCTGCTCCGAGGCTCTTCAGAGACCGCGGTG 127  
113 GAAATACAGATCATATGATTTGATTTCTGAATATGCTACAAACACTGCGTAG 172  
128 AGCAGCGCTCCCTATTTACCGCGCTTCTGAGACCTTCCACCTGAAGCGCTG 187  
173 ATACGATCATGATTTATTTATGATTTAAGATGATTTCCAGCATTTGAATTTAG 232  
188 CCGCGGCTTCTCTGAAAGGTCTTGAAGGTCTGCAAGGGGCGGCTACTACCGCGG 247  
233 CGAAGCGCGCATGATTAAGTTGTAAGCTTGGGAAGGTTAGTTATTCACGAG 292  
248 C-----GGAACACTCCAGCGCGCTGCGCGGAGGTGAGAGCTTCCC 292  
293 CGCTAATTAAAGTGGCAGCGCAAGATTTGTAAGATTTGGGAAATGCTTG 352  
293 CGAGCTTCGCGAGCTTCGCGGCGCTCTGCTCTGCTCGGCGCTTACACCGCGCGGTG 352  
353 ACACATCGAAGATTTTGGAGTTTAAAGAAATCGGCCCTTATACGCTGTGCGATTG 412  
353 CTTTCATCGCTTGGGGAAGCGGTGCGCGGCTGACGGAACGTCCGAGGCTCTCT 412  
413 GCAATATGCTTAAATCTCCGCAACAGCAATTTGACGATGATGCGGTGATGTA 472  
413 CCGCGCTTCTG 424  
473 GCCGTTATTG 484

## RESULT 14

US-09-385-028-23  
Sequence 23, Application US/09385028  
Patent No. 6232106  
GENERAL INFORMATION:  
APPLICANT: Susan E. Jensen  
APPLICANT: Kwame A. Aidoo  
APPLICANT: Ashish S. Paraskar  
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic  
Patent No. 6232106  
TITLE OF INVENTION: Acid Biosynthesis  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC

STREET: The Jenifer Building, 400 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/385,028  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/790,462  
FILING DATE: 29-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: D. Douglas Price  
REGISTRATION NUMBER: 24,514  
REFERENCE/DOCKET NUMBER: 1418/P57452US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 638-6666  
TELEFAX: (202) 39305350  
TELEX: RCA 248593 IDBA UR  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1227 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-385-028-23

Query Match 6.4%; Score 62.2; DB 3; Length 1227;  
Best Local Similarity 44.4%; Pred. No. 0.00055;  
Matches 345; Conservative 0; Mismatches 423; Indels 9; Gaps 2;

157 CTGAGACGCTTTCCACCTGAAAGCCCTGGCCGCGCTTCCCTGGAAGAGTCTTAGG 216  
88 CTGGCCGGGCTGGGTCCAGAAAGCCGAGCCGGGTGACGCTGTGGACGACGACCGAG 147  
217 GCTGCGAAGGGGGGGGCTAATACCGCGGGGGGGAACACTCCACCGCTGGCGCGAAGC 276  
148 GTGTGGCTGTGACCTTGACGCGCGGGGGCCCGGCTCTGGGCGACCGCGCTTACCC 207  
277 GTGAGAGAGCTTCCCGGAGCTTCCGCGAGCTTCCGCGGCTTCTGTGCTCGGCGCTTAC 336  
208 GCGGTACAGAGCCGCGCGCTTCCGAGTGTGACCCGCACTCCCACTGGTGGCGGCGC 267  
337 ACCGCGCGCGGGGTGCTTCATCGCTTCGGGGAGCGGGGTGGCGCGGTGACGG---- 392  
268 AACCCGGAATCGGCGTGTTCATCCGATGACGACCCGAGCACTCCGCGCTGCGCTCG 327  
393 --GAAGTCCGAGGGTCTCTCCGCGCTTCCGCGCGCGGAAAGCCCAAGAGAGAGAG 450  
328 ATGCTACCCGGAGCTTCTCTGCGCGCGCGCGAGCGCTGCGCGCGCGGTGCGGAG 387  
451 CTTTGGCCCTCGCCAGGAGCTCTCCCGAGGGGCTGAGCCCGGGGGTGTGAAACAG 510  
388 CTGCTGAGACGAGTCTCTGGCGGGGCTGTGTAAGGGGAGGCGGCTGCACTGTGCGC 447  
511 GCCCTCATGAGTCTGGGGGCGACGCTGTGCTGCGAAGAGCGCGCGGTGGCGGCGCTGC 570  
448 GGACTACGATCCCGGTGCGCTCGCGGGTATCACTCTCTTCCGCGCGCGGTGACGAG 507  
571 CCCCTAGGGGCTTCTGCGCGGGGGAAGAGGCCCGGGGCGTAACCCCGCGCCAGAGAG 630  
508 CGCGGGAGTTCAATCAGAGACCGCAGCGGGTCTTATGACACCGCGGCTACACCCGAG 567  
631 CGCGGGGAGAGAGAGAGCGCTGTGCTGTGCTGTCTCTCGGCGGAGAGGGGGTGCAC 690  
568 CAGGTGCGCAAGGCGCGGGAGCAACTGACGCGTATCTGTGCGGAGGCTGTGTGAGAGCGG 627

691 CTGAAAGCTTGAAGGGGCGCTTCCAGGCGCTCTAAGGGGCGTCCCTTTCCCGCTGAG 750  
628 ATCGAAGAACCCGGGACCGACCTGATCAGCGGCTCTCATTCAGCAGGTGCGGCGGG 687  
751 GAGCTCCCGGGCGGAGAGCGGCTTTCGGGTTAGTCTAAGGCCCTTAGCGAGGTGCGC 810  
688 CATCTGCGGCTCAGAGAGATGTCTCCGATGTGCGGCTGTGCTGTGTGCGCGTCAAGC 747  
811 CAGCCCTCACCACCGAGGCTTTCGGTGTGAGGTGCGGGGCGCTTTGGAGAGGGAG 870  
748 ACCACACACGAGCGAGGCTGAGCTGTCTGACCTGTCTACCGACCGAGACTGAGCC 807  
871 GGGAGAGACCCCTTGAAGAGGCC--CTACCAAGCTCATGAGAGGTGCTCCGC 924  
808 GGGGCTTACCGAGGACCCGCGCTGTGCTGCGCAAGCGGCTGAGAGACTGTCTGCGC 864

RESULT 15  
US-09-726-614-23  
Sequence 23, Application US/09726614  
Patent No. 6514735

GENERAL INFORMATION:  
APPLICANT: Susan E. Jensen  
APPLICANT: Kwamena A Aidoo  
APPLICANT: Aehish S. Paraskar  
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic  
Patent No. 6514735  
TITLE OF INVENTION: Acid Biosynthesis  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESS: JACOBSON, PRICE, HOLMAN & STERN, PLLC  
STREET: The Jenifer Building, 400 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/726,614  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/790,462  
FILING DATE: 29-JAN-1997  
ATTORNEY/AGENT INFORMATION:

NAME: D. Douglas Price  
REGISTRATION NUMBER: 24,514  
REFERENCE/DOCKET NUMBER: 1418/P57452US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 39305350  
TELEFAX: (202) 39305350  
TELEX: RCA 248593 IDBA UR  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1227 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-726-614-23

Query Match 6.4%; Score 62.2; DB 4; Length 1227;  
Best Local Similarity 44.4%; Pred. No. 0.00055;  
Matches 345; Conservative 0; Mismatches 423; Indels 9; Gaps 2;

157 CTGAGACGCTTTCCACCTGAAAGCCCTGGCCGCGCTTCCCTGGAAGAGTCTTAGG 216  
88 CTGGCCGGGCTGGGTCCAGAAAGCCGAGCCGGGTGACGCTGTGGAGACGACCGAG 147

OY	217	TTCTGGCAGGCGGCGGCTACTACCGCGGGGGGAAACACTTCCACCGCTGTGGCCGAAC	276
Db	148	GTGTGGCTGTGTGACTCTCGACCGCCGGGGCCCGAGCCGTCTGTGGGCAACCGCGCTTCAAC	207
OY	277	GTGAGAGAGCTTCCCCCGAGCTTCCGCGCATTCCTTGGGGGAGCTTCTGTGTCTGGGAGCTTAC	336
Db	208	GCGGTGACGAGCGCGCCGCCGCTTCCCGAATGTGTGACCCCGCACTTCCAACTGTGTGCGCGCC	267
OY	337	ACCAGGCGGCGGCGTGGAGCTTCATGCGCTTTCGGGGAGCGGGTGGCGGCGGTGACGG----	392
Db	268	AAACCGGAGTGGCGGTCGTTTCATCCCGCATGAGCAACCCGACGACTCCGCGGTGCGCTCG	327
OY	393	--GAACTGTCCGAGGGTCTCTTCCCGGCTTTCGCCCGGAAAAAGCCCAAGAGAAAGAG	450
Db	328	ATGTCTACCCCGGAGCTTCTCTGGCCCGCCGCGCGCATGGAGCCCTGCGCCCTCCGCTGCGGGAG	387
OY	451	CTTTTCCGCTTCGCCAGGGGCTCTCCGCCGAGGGCGTGTGAGCCCGGGGGGTGTGAACAG	510
Db	388	CTGCTGTGACGAGATCTTGGGCGGGCTGTGTGAAGGGGAGACGGCCGGTCCACTGTGTGCC	447
OY	511	GCCCTCATGAGAGCTGTGGGCGCAAGGTCTGTGCTGCGCAAAAGGACCCGTTGGAGGAGCTGCG	570
Db	448	GGACTGAGAGATCCCGGTGCTCTGGGGGATCATCCCTGTCTTTCGGCGCCGGTATCGAC	507
OY	571	CCCTTAGGGGCTTCTGTGCGGGGAGAGAGGCCCCCGGGCGCTACCCGCGCCTCAGAAAG	630
Db	508	CGCGGGAGTTTATCGAGGACCGGACGCGGATCTTATGACCGCGGCTTACACCCGGAG	567
OY	631	CGCCGGCGGAGAGAGAGCGCTCGTGCCTCTGTCTCTCTGGGCGGAAAGGGGGGTGCAC	690
Db	568	CAGGTGCGCAAGGCCCGGAGCAAACTCGACCGCTATCTGCGGGAAGTGTGTGAGAGCGG	627
OY	691	CTGGAAGAGCTTGAAGGGGCTTTCAGGGGCTCTACGGGGTCCCCCTCTTTCCTCCCTGAG	750
Db	628	ATCGAAGAACCGGGGACCGGACTTGATCAGCCGGGCTGTATGACCAAGGTGCGGCGGG	687
OY	751	GAGCTTCCCGGCGGAGCGGCTTTCGGGGTGTAGGCTTAGGCCCCCTTAGCGAGGTGCG	810
Db	688	CATCTGGGGGTGAGAGATGTGTCCGATGTGCGGGCTGTCTGTGTGGCCGGTACAGG	747
OY	811	CACGCTTCACTCCACCGAGGCTTGTGCGTGAAGGTGCGGGGGGCCCCCTTGTGGAGGGGAG	870
Db	748	ACCAACCAACAGCCAGGCGGAGCTGTGAGCTGTACGCTCTCAACCAACCCGGAGCTGGCC	807
OY	871	GGGAGAGACCCCTTGAGAGAGGCC---CTAACCAAGCTCATGAGAGAGGTGCTCGCG	924
Db	808	GGCGGCTTACCGAGAACCCGGGCTCTGTGCGCCCAAGCGGTGAGGAGTGTGCTGCTCC	864

Search completed: March 4, 2004, 05:48:50  
Job time : 97 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: March 4, 2004, 03:03:25 ; Search time 358 Seconds  
(without alignments)  
11569.813 Million cell updates/sec

Title: US-09-938-901-1

Perfect score: 975  
Sequence: 1 gtggaagccgcgcgaagc.....tagtcccccccgagcga 975

Scoring table: IDENTITY NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : N\_Geneseq\_29Jan04:\*  
Listing first 45 summaries

1: N\_Geneseq19808:\*  
2: N\_Geneseq19908:\*  
3: N\_Geneseq20008:\*  
4: N\_Geneseq20018:\*  
5: N\_Geneseq20028:\*  
6: N\_Geneseq20038:\*  
7: N\_Geneseq20048:\*  
8: N\_Geneseq20058:\*  
9: N\_Geneseq20068:\*  
10: N\_Geneseq20078:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	975	100.0	975	7	ABZ22143
2	104	10.7	1608	9	ABZ22143 Thermus t
3	104	10.7	1608	9	ADE36258 Human MYH
4	104	10.7	1608	9	ADE36259 Human MYH
5	104	10.7	1854	7	ABV72438
6	104	10.7	1858	2	ABT8194
7	102	10.5	1878	2	AACT8090 Human mis
8	102	10.5	1856	2	AAV35701 CDNA enco
9	95.6	9.8	349980	3	AAE36257 Human MYH
10	95.4	9.8	110000	3	AAE36257 Human MYH
11	92.8	9.5	528	6	ABK74292 Bacillus
12	89.2	9.1	1047	7	ABZ40024 N. gonorr
13	87.2	8.9	2293	4	AAE63242 Escherich
14	86.6	8.9	912	7	ABZ79956 Mycobacte
15	86.6	8.9	1312	7	ABZ79955 Mycobacte
16	86.6	8.9	110000	4	AAI99682_40
17	86.6	8.9	110000	4	AAI99683_40
18	83.6	8.6	23532	3	AAE81455 N. mening
19	82	8.4	1173	7	ABX06945 S. pneumo
20	82	8.4	9909	7	AAV52145 Streptococ
21	82	8.4	110000	7	ABX56454_10
22	81.2	8.3	968	6	AAE25855 Micrococc
23	81	8.3	960	2	AAK34829 S. pneumo

C	24	81	8.3	2077	2	AAZ96352	AAZ96352 S. pneumo
C	25	81	8.3	2077	2	AAV37388	AAV37388 Streptococ
C	26	80.4	8.2	1176	2	AAK34828	AAK34828 S. pneumo
C	27	80.2	8.2	1038	7	ACF69235	ACF69235 Photornab
C	28	80.2	8.2	110000	7	ACF67367_21	ACF67367 Photornab
C	29	80.2	8.2	243072	7	ACF65382	ACF65382 Photornab
C	30	78.6	8.1	1206	8	ADB10739	ADB10739 Alloiococ
C	31	78.6	8.1	1206	8	ADB10741	ADB10741 Alloiococ
C	32	78.6	8.1	1206	8	ADB10743	ADB10743 Alloiococ
C	33	78.6	8.1	110000	8	ADB12064_11	ADB12064_11 Alloiococ
C	34	75	7.7	114955	2	AAK53491	AAK53491 Human ade
C	35	73.4	7.5	66788	4	AAK59515	AAK59515 Propionib
C	36	73.4	7.5	66788	7	ACF64444	ACF64444 Propionib
C	37	72.6	7.4	349980	6	ABO81847	ABO81847 Bifidobac
C	38	71.8	7.4	1412	3	AACT8090	AACT8090 Human ORF
C	39	71.8	7.4	1412	6	ABN22977	ABN22977 Human ORF
C	40	70.6	7.2	114955	2	AAK53491	AAK53491 Human ade
C	41	68.2	7.0	991	6	ABO68775	ABO68775 Listeria
C	42	68.2	7.0	3155	6	ABO70783	ABO70783 Listeria
C	43	66	6.8	35133	4	ABL50991	ABL50991 Thermus c
C	44	66	6.8	35134	4	ABL50990	ABL50990 Thermus c
C	45	65.6	6.7	2000	7	ADA71938	ADA71938 Rice gene

## ALIGNMENTS

RESULT 1	ABZ22143	standard; DNA; 975 BP.
ID	ABZ22143	
XX	ABZ22143;	
AC	12-MAR-2003	(first entry)
DT		
XX	Thermus thermophilus DNA repair enzyme MutY encoding DNA SEQ ID NO:1.	
DE	Thermus thermophilus; DNA repair enzyme; MutY; RecJ; TRCF;	
XX	biochemistry; molecular biology; research; gene; ds.	
KW	Thermus thermophilus.	
XX	Key	Location/Qualifiers
OS	Key	1..975
FT	CDs	/*tag= a
FT		/partial
FT		/product= "MutY"
FT		/note= "DNA repair enzyme; no start or stop codons given"
XX	JP2002247985-A.	
XX	03-SEP-2002.	
XX	23-FEB-2001; 2001JP-00047762.	
XX	23-FEB-2001; 2001JP-00047762.	
XX	(RIKA) RIKAGAKU KENKYUSHO.	
XX	WPI: 2003-078924/08.	
XX	P-PSDB; ABP56413.	
XX	A DNA repair enzyme gene, a protein, a recombinant vector, a	
XX	transformant, preparation of DNA repair enzyme, repairing the error	
XX	sequence of a DNA, and prevention of error synthesis of a DNA sequence.	
XX	Claim 3; Page 13-14; 41pp; Japanese.	
XX	ABZ22143 to ABZ22146 encode the Thermus thermophilus DNA repair enzymes	
XX	MutY, RecJ, RecF, and TRCF given in ABP56413 to ABP56416. The enzymes can	
XX	be used as research reagents for biochemistry and molecular biology	
XX	Sequence 975 BP; 126 A; 347 C; 354 G; 148 T; 0 U; 0 Other;	

```
Query Match      100.0%; Score 975; DB 7; Length 975;
Best Local Similarity 100.0%; Pred. No. 4, Se-160;
Matches 975; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GTGAGAGCTGCGGAGAAAGCCCTCTGCGCTGTATCCGGGAAAAAGCCCGCCCTCCCC
DB      1 GTGAGAGCTGCGGAGAAAGCCCTCTGCGCTGTATCCGGGAAAAAGCCCGCCCTCCCC
QY      61 TGGCGGGGAGAGAGAACCTTACCGCGTCTGTGCTCCGAGGCTCTTCTGACAGACACC
DB      61 TGGCGGGGAGAGAGAACCTTACCGCGTCTGTGCTCCGAGGCTCTTCTGACAGACACC
QY      121 CGGGTGGACAGAGCCCTCCCTTTATACCGCGCGCTTTCTTGGAGGCTTTCCACCCCTGAAG
DB      121 CGGGTGGACAGAGCCCTCCCTTTATACCGCGCGCTTTCTTGGAGGCTTTCCACCCCTGAAG
QY      181 GCGCTGCGCGGCTTCCCTGTGAGAGGTCCTTAGGGTCTGGCAGAGGGGCGGGCTACTAC
DB      181 GCGCTGCGCGGCTTCCCTGTGAGAGGTCCTTAGGGTCTGGCAGAGGGGCGGGCTACTAC
QY      241 CGGCGGGGAGAACACTTCCACCGCTGCGCGGAGGCTGTGAGAGGCTTCCCGAGCTTC
DB      241 CGGCGGGGAGAACACTTCCACCGCTGCGCGGAGGCTGTGAGAGGCTTCCCGAGCTTC
QY      301 GCGGAGCTTCGGGGGCTTCCTGCTCGGCTTACACCGCGCGGCGGGGCTGCGCTCAGC
DB      301 GCGGAGCTTCGGGGGCTTCCTGCTCGGCTTACACCGCGCGGCGGGGCTGCGCTCAGC
QY      361 GCGCTTGGGGAGAGGGTGGCGCGCGGTGAGACGGAGAGTCCGAGAGGCTCTCTCCGCTC
DB      361 GCGCTTGGGGAGAGGGTGGCGCGCGGTGAGACGGAGAGTCCGAGAGGCTCTCTCCGCTC
QY      421 TTGCGCCGAGAAAGCCCAAGAGAGAGCTTTTGCCCTGCGCCAGGAGCTCTCTCC
DB      421 TTGCGCCGAGAAAGCCCAAGAGAGAGCTTTTGCCCTGCGCCAGGAGCTCTCTCC
QY      481 GAGGGGCTGAGACCGGGGGGTGTGAGACAGAGCCCTATAGAGTCTGGGGCCACAGGCTAC
DB      481 GAGGGGCTGAGACCGGGGGGTGTGAGACAGAGCCCTATAGAGTCTGGGGCCACAGGCTAC
QY      541 CTGCGGAAAGCGCCCGTGGCGGGGCTGCGCCCTTACGGGAGCTTCTGCGGGGAGAGAG
DB      541 CTGCGGAAAGCGCCCGTGGCGGGGCTGCGCCCTTACGGGAGCTTCTGCGGGGAGAGAG
QY      601 GCGCCCGGGGCGCTACCCCGCGCCAGAGAGCGCGGGCGAAGAGAGCGCTCTGCGCC
DB      601 GCGCCCGGGGCGCTACCCCGCGCCAGAGAGCGCGGGCGAAGAGAGCGCTCTGCGCC
QY      661 CTGCTCTCTGCGGGGAGAGGGGGGTGACCTGGAAGAGCTTGAAGGGGGGCTTCCAGGGC
DB      661 CTGCTCTCTGCGGGGAGAGGGGGGTGACCTGGAAGAGCTTGAAGGGGGGCTTCCAGGGC
QY      721 CTCTACGCGGCTCCCTCTTTTCCCTGAGAGGCTTCCCGGGGCGGAGCGGCTTCCGG
DB      721 CTCTACGCGGCTCCCTCTTTTCCCTGAGAGGCTTCCCGGGGCGGAGCGGCTTCCGG
QY      781 GTGAGGTCTAAGCCCTTAGGCGAGGTGCGCCACGCTTACCCACCGAGGCTTGGCGTG
DB      781 GTGAGGTCTAAGCCCTTAGGCGAGGTGCGCCACGCTTACCCACCGAGGCTTGGCGTG
QY      841 GAGGTGCGGGGGGCGCTTTGGAGAGGGGAGAGAGCCCTGGAAGAGGGCGCCCTACCC
DB      841 GAGGTGCGGGGGGCGCTTTGGAGAGGGGAGAGAGAGCCCTGGAAGAGGGCGCCCTACCC
QY      901 AAGCTCATGAGAAAGGTGTCTCGCAAGCGCTTCCCTCTGCTCATGCGGGCGTAGTC
DB      901 AAGCTCATGAGAAAGGTGTCTCGCAAGCGCTTCCCTCTGCTCATGCGGGCGTAGTC
QY      961 CCCCTCCGAGCGCA 975
DB      961 CCCCTCCGAGCGCA 975
```

```
RESULT 2
ADE36258
ID ADE36258 standard; cDNA, 1608 BP.
XX
AC ADE36258;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human MYH (hMYH) G1396T mutant cDNA.
XX
KW gene; ss; mutant; human; screening method; hMYH; base excision repair;
KW BER; APC; familial adenomatous polyposis; FAP;
KW multiple colorectal adenoma; carcinoma; bowel cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1608
FT /tag= a
FT /product= "Human MYH mutant protein (E466X)"
FT /replace(1396..9)
FT /tag= b
FT /note= "Base T in mutant replaces wild type base G"
XX
PN MO2003014390-A2.
XX
PD 20-FEB-2003.
XX
PF 02-AUG-2002; 2002MO-GB003591.
XX
PR 03-AUG-2001; 2001GB-00018995.
XX
PA (UYWA-) UNIV WALES COLLEGE OF MEDICINE.
XX
PI Sampson JR, Cheadle JP;
XX
DR WPI: 2003-256601/25.
DR P-PSDB: ADE36261.
XX
PT Screening, diagnostic and therapeutic methods in individuals with
PT predisposition towards having a cancer, such as colon cancer, using base
PT excision repair pathway or hMYH genes.
XX
PS Claim 9; SEQ ID NO 3; 66bp; English.
XX
CC This invention relates to a novel screening method for identifying an
CC individual having a predisposition towards a cancer. Specifically, it
CC refers to obtaining a test sample, preferably comprising the hMYH gene
CC that occurs in the base excision repair (BER) pathway, and comparing this
CC nucleic acid molecule to the corresponding region of the wild type
CC sequence. This BER pathway gene, hMYH, acts to protect against G:C to T:A
CC transverse mutations in a cancer marker gene such as APC that is seen in
CC familial adenomatous polyposis (FAP). As such, mutations identified in
CC hMYH are associated with the onset multiple colorectal adenomas and
CC carcinoma. The present invention describes a screening method for
CC individuals that works to identify differences comprising any one of
CC G387D, Y165C, E466X or Y90X variations in hMYH, this signifies a cancer
CC predisposition, particularly for bowel cancer. This polynucleotide
CC sequence is the mutant G1396T hMYH cDNA that encodes a E466X mutation in
CC the encoded polypeptide of the invention.
XX
SQ Sequence 1608 BP; 337 A; 478 C; 502 G; 291 T; 0 U; 0 Other;
XX

Query Match      10.7%; Score 104; DB 9; Length 1608;
Best Local Similarity 54.0%; Pred. No. 2,66-09;
Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;
```

```
QY      69 GGAGAAGAACCTTACCGCGCTCTGCTCCGAGTCTTGTGACGACAGACCCGGGTGGA
DB      327 GGACAGGCGGCGCATATGCTGTGTGCTCAGAGTCTATGCTGACGACGACCGATTGC
QY      129 GCAGGCCCTCCCTATTACCGCGCGCTTCTGAGCGGCTTCCACCTGAAGGCCCTGGC
DB      129 GCAGGCCCTCCCTATTACCGCGCGCTTCTGAGCGGCTTCCACCTGAAGGCCCTGGC
```

Db 387 CACTGTATCACTACTATACCGGATGATGACAAAGTGGCTTACACTGACAGACCTTGGC 446  
Qy 189 GCGGGCTTCCCTGAGAGAGTCTTAAAGGTCTGGCAGGGGGCGGCTACTACCGGGGC 248  
Db 447 CAGTGTCTCCCTGAGAGAGTGAATCACTGTGGCTGGCTGGGCTACTATTTCTCGTGG 506  
Qy 249 GGAACACTCCACCGGCTGGGCGGAGGCTTGGAGAGCTTCCCGAGCTTC----- 300  
Db 507 CCGGCGGCTGCAGAGAGGAGCTCGAAGGTGTGTAGAGAGCTAAGGGGCGCATGCGACG 566  
Qy 301 -----GCCGAGCTTGGGGGGCTTCTGTGTGGGCTTACACCGGGCGGCGGT 350  
Db 567 TACAGCAGAGACTCTGACAGCTTCTGCTGGCTGGGGCGCTACACAGCTGGGGCCAT 626  
Qy 351 GGCCTTCATCGCTTTCGGGGAGCGGGTGGCGCGATGAGCGGAACGTCCGAGAGTCTT 410  
Db 627 TGCCTTATGCGCTTTTGGCCAGGACCGGTGTGTGGATGGCAACGTAGCAGCGGGTCT 686  
Qy 411 CTCCCGCTTTCGCCC--GGAAAAGCCCCAAGAGAAAGAGCTTTTGGCCCTTGGCCAG 468  
Db 687 GTGCCGTGTCCGACCACTGTGTCTGATCCAGACAGCACCCTTGTTCCTCAGCAGCTCTG 746  
Qy 469 GGCCTCTCCCGAGGGGGGTGACCCG-----GGGGTGTGAACCAAGCGCTTCAT 518  
Db 747 GGGCTTAGCCAGAGCTGTGTGACCCAGCCCGCCAGAGATTTCAACCAAGACCAT 806  
Qy 519 GAGAGCTGGGGGCAAGGCTTGTCTGCGGAAAGCGCCCGTTCGGGGGCTGCCCCTAGG 578  
Db 807 GAGAGCTAGGGGGCCACAGTGTGTACCCACAGCGCCCACTGTGACAGCCAGTCCCTGTGA 866  
Qy 579 GGCCTTGTCCGGG 592  
Db 867 GAGCTGTGCCGG 880

RESULT 3  
ADE36259  
ID ADE36259 standard; cDNA; 1608 BP.  
AC ADE36259;  
XX  
DT 29-JAN-2004 (first entry)  
DE Human MYH (hMYH) C270A mutant cDNA.  
XX  
KW gene; ss; mutant; human; screening method; hMYH; base excision repair;  
KW BBR; APC; familial adenomatous polyposis; FAP;  
KW multiple colorectal adenoma; carcinoma; bowel cancer.  
XX  
OS Homo sapiens.  
XX  
FH Key 1.1608  
FT CDS location/Qualifiers  
FT /\*tag= a  
FT /product= "Mutant MYH protein (Y90X)"  
FT replace(270,c)  
FT /\*tag= b  
FT /note= "Base A in the mutant replaces wild type base C"  
XX  
XX  
XX W02003014390-A2.  
XX  
XX 20-FEB-2003.  
XX  
XX 02-AUG-2002; 2002WO-GB003591.  
XX  
XX 03-AUG-2001; 2001GB-00018995.  
XX  
XX (UYMA-) UNIV WALES COLLEGE OF MEDICINE.  
XX  
XX Sampson JR, Cheadle JP;  
XX  
XX WPI; 2003-256601/25.  
XX  
XX

DR P-PSDB; ADE36262.  
XX  
XX Screening, diagnostic and therapeutic methods in individuals with  
FT predilection towards having a cancer, such as colon cancer, using base  
FT excision repair pathway or hMYH genes.  
XX  
PS Claim 10; SEQ ID NO 4; 66pp; English.  
XX  
XX This invention relates to a novel screening method for identifying an  
CC individual having a predilection towards a cancer. Specifically, it  
CC refers to obtaining a test sample, preferably comprising the hMYH gene  
CC that occurs in the base excision repair (BER) pathway, and comparing this  
CC nucleic acid molecule to the corresponding region of the wild type  
CC sequence. This BER pathway gene, hMYH, acts to protect against G:C to T:A  
CC transverse mutations in a cancer marker gene such as APC that is seen in  
CC familial adenomatous polyposis (FAP). As such, mutations identified in  
CC hMYH are associated with the onset multiple colorectal adenomas and  
CC carcinoma. The present invention describes a screening method for  
CC individuals that works to identify differences comprising any one of  
CC G382D, Y165C, E466X or Y90X variations in hMYH, this signifies a cancer  
CC predilection, particularly for bowel cancer. This polynucleotide  
CC sequence is the mutant C270A hMYH cDNA that encodes a Y90X mutation in  
CC the encoded polypeptide of the invention.  
XX  
SQ Sequence 1608 BP; 338 A; 477 C; 503 G; 290 T; 0 U; 0 Other;  
Query Match 10.7%; Score 104; DB 9; Length 1608;  
Best Local Similarity 54.0%; Pred. No. 2.6e-09;  
Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;  
Qy 69 GGAAGAAGACCTTACCGCGCTCTGCTCCGAGGCTCTTGTGAGACAGACCGGGTGA 128  
Db 327 GAGCAGGCGGGCATATGCTGTGTGTCTAGAGGTATGTGACAGACCAAGTTTC 386  
Qy 129 GCAAGCCCTCCCTATTACCGCGCTTTCGAGCGCTTTCACACCTGAAAGGCTTGC 188  
Db 387 CACTGTATCACTACTATACCGGATGATGACAAAGTGGCTTACCTGACAGACCTTGC 446  
Qy 189 GCGGCTTCCCTGAGAGAGTCTTAAAGTGTGCAAGGGGGCGGCTTACTACCGCGGGC 248  
Db 447 CAGTGTCTCCCTGAGAGAGTGAATCACTGTGGCTGGCTGTGCTACTATTTCTGTGG 506  
Qy 249 GGAACACTCCACCGGCTGGGCGGAGGTGAGAGAGCTTCCCGAGCTTC----- 300  
Db 507 CCGGCGGCTGCAGAGAGGAGTTCGAAAGTGTGTAGAGAGTTAGGGGCGCATGCCACG 566  
Qy 301 -----GCCGAGCTTGGGGGGCTTCTGTGTGGGCTTACACCGGGCGGCGGT 350  
Db 567 TACAGCAGAGACTCTGACAGCTTCTGCTGGCTGGGGCGCTACACAGCTGGGGCCAT 626  
Qy 351 GGCCTTCATCGCTTTCGGGAGCGGGTGGCGCGATGAGCGGAACGTCCGAGAGTCTT 410  
Db 627 TGCCTTATGCGCTTTTGGCCAGGACCGGTGTGTGGATGGCAACGTAGCAGCGGGTCT 686  
Qy 411 CTCCCGCTTTCGCCC--GGAAAAGCCCCAAGAGAAAGAGCTTTTGGCCCTTGGCCAG 468  
Db 687 GTGCCGTGTCCGACCACTGTGTCTGATCCAGACAGCACCCTTGTTCCTCAGCAGCTCTG 746  
Qy 469 GGCCTCTCCCGAGGGGCTGACCCG-----GGGGTGTGAACCAAGCGCTTCAT 518  
Db 747 GGGCTTAGCCAGAGCTGTGTGACCCAGCCCGCCAGAGATTTCAACCAAGACCAT 806  
Qy 519 GAGAGCTGGGGGCAAGGCTTGTCTGCGGAAAGCGCCCGTTCGGGGGCTGCCCCTAGG 578  
Db 807 GAGAGCTAGGGGGCCACAGTGTGTACCCACAGCGCCCACTGTGACAGCAGTCCCTGTGA 866  
Qy 579 GGCCTTGTCCGGG 592  
Db 867 GAGCTGTGCCGG 880

RESULT 4  
ABV72438

ID ABV72438 standard; DNA; 1854 BP.  
 AC ABV72438;  
 XX  
 DT 29-JAN-2003 (first entry)  
 XX  
 DE Nucleotide sequence of adenine mismatch glycosylase.  
 XX  
 KM Human; cancer; DNA repair enzyme; DNA damage preventing enzyme;  
 XX adenine mismatch glycosylase; ss.  
 OS Homo sapiens.  
 PN W0200276280-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-IL000231.  
 XX  
 PR 23-MAR-2001; 2001US-00815015.  
 PR 09-JUL-2001; 2001US-030338P.  
 XX  
 PA (YEDA ) YEDA RES & DEV CO LTD.  
 XX  
 PI Livneh Z, Paz-Elizur T, Blumenstein S;  
 DR WPI; 2003-029894/02.  
 XX  
 PT Determining the risk to develop cancer for prevention, early detection  
 PT and prognosis, comprises determining a level of parameter indicative of a  
 PT level of activity of a DNA repair/damage preventing enzyme in a tissue of  
 PT the subject.  
 XX  
 PS Disclosure; Page 86-87; 91pp; English.  
 XX  
 CC The specification describes a method for determining the risk of a  
 CC subject to develop cancer. The method comprises determining a level of  
 CC parameter indicative of a level of activity of a DNA repair/damage  
 CC preventing enzyme in a tissue of the subject, and determining the risk of  
 CC the subject to develop the cancer according to the level. The method is  
 CC useful for determining risk of developing cancer, evaluating the  
 CC effectiveness and dosage of a cancer therapy, and determining the  
 CC presence of correlation or non-correlation between activity of at least  
 CC one DNA repair enzyme and a cancer for prevention, early detection and  
 CC prognosis of cancer. ABV72432-43 encode DNA repair/damage preventing  
 CC enzymes, whose level or activity may be determined in the method of the  
 CC invention  
 CC  
 SQ Sequence 1854 BP; 397 A; 546 C; 568 G; 343 T; 0 U; 0 Other;  
 Query Match 10.7%; Score 104; DB 7; Length 1854;  
 Best Local Similarity 54.0%; Pred. No. 2.5e-09;  
 Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;  
 QY 69 GGAGAAGACCCCTTACCGGCTCTGTCGCCAGGTCCTTCTGCGACGACCCCGGTGA 128  
 DB 494 GGACAGCGGGCATATGCTGTGTGCTCAAGGTCATGTCGACGACCCGAGTTGC 553  
 QY 129 GGAAGCCCTCCCTATTACCGCGCTTCTGAGGCTTCCACCCGAAAGCCCTGGC 188  
 DB 554 CACTGTGATCACTACTATACCGGATGATGACGAAGTGGCTACACGACGACTGGC 613  
 QY 189 CGCGGTTTCCCTGGAAGAGCTCTTAGGCTCTGCGAGGGGCGGCTTACCTACCGCGGAC 248  
 DB 614 CAGTGTCTTCTGAGAGAGTGAATCACTTGGGCTGGCTGGGCTACTATTCTCGTGG 673  
 QY 249 GGAAACCTCCACCGCCCTGGCCGGAAGGTGAGAGGCTTCCCGAGCTTC----- 300  
 DB 674 CGGCGGCTGCGAGGAGGCTCGGAAGCTGTAGAGAGCTTAAGGGGCCACATGCCACG 733  
 QY 301 -----GCCGAGCTTTCGGGGGCTTCTGTGCTCGGGCTTACACCGCGCGCGGT 350  
 DB 734 TACAGCAGAGACCCCTGACGAGCTCTGCTGCGGTGGGGCGCTACACAGCTGGGGCAT 793

QY 351 GGCCTCATTCGCTTTCGGGAGCGGGTGGCGCGGTGACGGGAAGTCCGAGGCTCT 410  
 DB 794 TGCCTATCGCTTTGGCCAGGCAACCGGTGTGTGATGGCAACGTAGCACGGGTCT 853  
 QY 411 CTCGCGCTCTTTCGGCC--GGGAAGCCCAAGAGAGAGCTTTTGGCCCTGGCCAG 468  
 DB 854 GTGCCCTGTCCAGAGCATTTGTGCTGATCCAGACACCCCTTTTCCAGACCTCTG 913  
 QY 469 GGCCTCTCTCCGAGGGCGGTGACCG-----GGGTGTGAACAGGCGCTTCAT 518  
 DB 914 GGATCTAGCCCAAGCTGTGSAACCCAGCCGGCCAGGAGATTTCACCAAGACCAT 973  
 QY 519 GGAGCTCGGGGCAAGGCTTCCTGCGCAACCGCCCGTTCGGGGCTGCCCCCTAGG 578  
 DB 974 GAGCTTAGGGGACAGTGTATCCCAAGCGGCCACTGTGCAAGCCAGTCCCTGTGA 1033  
 QY 579 GGCCTTTCGCGGG 592  
 DB 1034 GAGCTGTGCGGG 1047

## RESULT 5

AA89194  
 ID AA89194 standard; cDNA; 1858 BP.  
 XX  
 AC AA89194;  
 XX  
 DT 27-MAR-1998 (first entry)  
 XX  
 DE Human mismatch repair MutY cDNA (hMYH gene).  
 XX  
 KM MutY; hMYH gene; mismatch repair; non-polyposis colon cancer;  
 KM xeroderma pigmentosum; gene therapy; diagnosis; human; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PH Key Location/Qualifiers  
 FT CDS 172..1731  
 FT /\*tag= a  
 FT variation 537  
 FT /\*tag= b  
 FT /\*note= "cytosine replaces guanine"  
 FT variation 852  
 FT /\*tag= c  
 FT /\*note= "cytosine replaces thymine"  
 FT variation 1286  
 FT /\*tag= d  
 FT /\*note= "cytosine replaces guanine"  
 XX  
 PN W09733903-A1.  
 XX  
 PD 18-SEP-1997.  
 XX  
 PF 11-MAR-1996; 96WO-US003239.  
 XX  
 PR 11-MAR-1996; 96WO-US003239.  
 XX  
 PA (SMIT ) SMITHKLINE BEECHAM CORP.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Wei Y;  
 XX  
 DR WPI; 1997-470811/43.  
 DR P-PsDB; AAW31912.  
 XX  
 PT DNA encoding human MutY protein - useful for detecting and treating  
 PT mismatches in DNA especially in non-polyposis colon cancer and xeroderma  
 PT pigmentosum.  
 XX  
 PS Claim 6; Page 37-39; 59pp; English.  
 XX  
 CC This cDNA clone codes for human MutY (see AAW31912), which is homologous



CC to the Escherichia coli MutY protein involved in the pathway that  
CC corrects A/G and A/C mismatches as well as adenines paired with 7,8-  
CC dihydro-8-oxo-deoxyguanine in mutated DNA. It was isolated from a cDNA  
CC library derived from human cerebellum tissue. The hmyh gene contains 15  
CC introns and is 7.1 kb long. Some naturally occurring allelic variants of  
CC the coding sequence have been identified. Polynucleotides at least 70%  
CC identical to the 1811 bp sequence, or complementary to such a sequence,  
CC and comprising at least 30 nucleotides, can be incorporated into a vector  
CC and used for production of recombinant polypeptides in host cells. The  
CC MutY polypeptide, or the polynucleotide encoding it, can be used e.g. to  
CC repair oxidative damage to DNA, to prevent mutations from oxidative  
CC lesions, to treat genetic diseases related to a mutated hmyh gene, e.g.  
CC xeroderma pigmentosum and neoplasia, and to diagnose an abnormal  
CC transformation or a susceptibility to abnormal transformation of cells,  
CC particularly in a non-polyploid colon cancer

XX Sequence 1858 BP; 401 A; 546 C; 567 G; 344 T; 0 U; 0 Other;

Query Match 10.7%; Score 104; DB 2; Length 1858;

Best Local Similarity 54.0%; Pred. No. 2.5e-09; Mismatches 225; Indels 30; Gaps 3;

69 GGAGAGAGACCTTACCGCGCTGCTCCGAGTCTCTTGGAGAGACCGCGGTGA 128  
498 GGACAGCGCGGCATATCTGTGTGCTTCAGAGTCACTGTGACAGACCAAGCTTGC 557  
129 GCAGGCGCTCCCTATTATCCCGCGCTTTGGAGCGCTTCCACCTGAAGCGCTTGC 188  
558 CACTGTATCAACTACTATACCGATGGATGACAGAGTGGCTACACTGACAGACCTGGC 617  
189 CGCGGCTTCCCTGGAAGAGTCTTAAAGGTGTGGCAGGGGGGGGCTACTACCGCGGCG 248  
618 CAGTCTTCCCTGAGAGAGTGAATCAACTCTGGGCTGGCTGAGTACTATCTCTGGG 677  
249 GGAACACTCCACCGCTGCGCGGAGAGAGTTCCTCCCGAGCTTC----- 300  
678 CCGCGGCTGAGAGAGAGTCTCGAAGGTGTAGAGAGTACAGGGGCCACATGCCACG 737  
301 -----GCCGAGCTTGGGGGGCTTCTGTCTCGGGGCTTACACCGGCGGCGGT 350  
738 TACAGAGAGACCTTGACAGAGCTCTGCTGGCGGGGGCTACACAGCTGGGGCCAT 797  
351 GGCTTCATGCTTCCGGGAGCGGGGTGGCGCGGTGAGACGGGAACGTCGGAGGCTCT 410  
798 TGCTCTATGCTGCTTGGCCAGGACCGGTGTGTGATGGCACTAGACAGCGGTGCT 857  
411 CTCGCGCTCTTGGCGCC--GGGAAAGCCCCAAGAGAGAGCTTTTGGCCCTGCGCCAG 468  
858 GTGCGGTGTGCGAGCCATTGTGTGTGATCCAGACGACCTTTTCCAGCAGCTTGTG 917  
469 GGCCTCTCCCGAGGGCGTGGACCG-----GGGGTGTGAAACCAAGGCTTCAT 518  
918 GGGGTATGAGCCAGAGCGTGTGACCCAGCGCGGCAAGAAATTTCACCAAGCAGCAT 977  
519 GGAAGTGGGGGACCGGTCTGCTGCGGAACCGGCGCGGCTTGGCGGGGCTTCCCTTAA 578  
978 GGAAGTGGGGGACCGGTGTGATCCCAAGCGCGCACTGTGACGAGCAGTGCCTGTGGA 1037  
579 GGCCTTGGCGCGG 592  
1038 GAGCTGTGCGGG 1051

RESULT 6  
ID AAC78090  
AAC78090 standard; cDNA; 1878 BP.

XX AAC78090;

DT 08-FEB-2001 (first entry)

XX Human cancer associated gene sequence SEQ ID NO:484.

KW Human; cancer associated gene; cancer antigen; detection; cancer;  
KW diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;  
KW antidiabetic; antistatic; antirheumatic; antithrombotic; antiviral;  
KW antidiabetic; antithyroid; antiallergic; antibacterial; cardiant;  
KW dermatological; neuroprotective; thrombolytic; coagulant; neoplastic;  
KW vasotrophic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
KW allergic reaction; graft versus host disease; organ rejection;  
KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
KW neurological disease; drug screening; ss.

OS Homo sapiens.

XX MO200055350-AL.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000MO-US005882.

XX 12-MAR-1999; 99US-0124270P.

XX (HUMA-) HUMAN GENOME SCT INC.

XX Rosen CA, Ruben SM;

XX WPI, 2000-58753/55.

XX P-PSDB; AAB43881.

XX Novel isolated nucleic acids comprising sequences encoding peptides  
XX useful for treating or diagnosing e.g. cancer.

XX Claim 1; Page 1016-1017; 2352pp; English.

XX AAC78090 to AAC78448 encode the human cancer associated proteins given in  
XX AAB43398 to AAB44239. The proteins can have activities based on the  
XX tissues and cells the genes are expressed in. Example of activities  
XX include: cytostatic; proliferative; vulnerability; immunomodulator;  
XX antidiabetic; antistatic; antirheumatic; antithrombotic;  
XX antidiabetic; antithyroid; antiallergic; antibacterial; antiviral;  
XX dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
XX neoplastic; vasotrophic; antipsoriatic; antiangiogenic. The  
XX polynucleotides and polypeptides can be used for preventing, treating or  
XX ameliorating medical conditions and diagnosing pathological conditions.  
XX Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
XX the present invention may be used to treat immune disorders by activating  
XX or inhibiting the proliferation, differentiation or mobilisation of  
XX immune cells, to treat disorders of haematopoietic cells, autoimmune  
XX disorders, allergic reactions, graft versus host disease and organ  
XX rejection, modulate haemostatic or thrombolytic activity, modulate  
XX inflammation, cancers, cardiovascular disorders, neurological disease and  
XX bacterial or viral infections. The peptides, nucleotides, antibodies,  
XX agonists and antagonists may be also be used in drug screens. AAC78449 to  
XX AAC78457 and AAB44240 represent sequences used in the exemplification of  
XX the present invention

XX Sequence 1878 BP; 411 A; 548 C; 566 G; 348 T; 0 U; 5 Other;

Query Match 10.7%; Score 104; DB 3; Length 1878;

Best Local Similarity 54.0%; Pred. No. 2.5e-09; Mismatches 225; Indels 30; Gaps 3;

69 GGAGAGAGACCTTACCGCGCTGCTCCGAGTCTCTTGGAGAGACCGCGGTGA 128  
509 GGACAGCGCGGCATATCTGTGTGCTTCAGAGTCACTGTGACAGACCAAGCTTGC 568  
129 GCAGGCGCTCCCTATTATCCCGCGCTTTGGAGCGCTTCCACCTGAAGCGCTTGC 188  
569 CACTGTATCAACTACTATACCGATGGATGACAGAGTGGCTTACCTGACAGACCTTGC 628  
189 CGCGGCTTCCCTGGAAGAGTCTTAAAGGTGTGGCAGGGGGGGGCTACTACCGCGGCG 248  
629 CAGTCTTCCCTGAGAGAGTGAATCAACTCTGGGCTGGCGCTGACTATCTCTGGG 688

```
QY 249 GGAACACTCCAGCGCTGGCCGAAAGCTGAGAGAGCTTCCCCGAGCTT----- 300
Db 689 CCGGCGGCTGCAAGAGAGAGCTGGAAGGTGAGAGAGCTAGGGGCCACATGCCACG 748
QY 301 -----GCCGAGCTTCGAGGAGCTTCTGCTCGGCTTACACGCGCGCGGT 350
Db 749 TACAGCAGAGAGCTGACAGAGCTCTGCTGGCGTGGGGCGCTACACAGCTGGGCGAT 808
QY 351 GGCCTTCATTCGCTTGGGAGAGGAGGTGGCGCGGTGAGAGAGAGAGCTCCGAGAGGTCT 410
Db 809 TGCCTTATTCGCTTGGCGAGAGCAACGATGTGTGATGCAAGTACAGAGGTGCT 868
QY 411 CTCGCCCTCTTTCGCCC--GGGAAAGCCCAAGAGAGAGCTTTCGCCCTCGCCAG 468
Db 869 GTGCCCTGTCCGAGCATTGTGTCTGATCCAGACAGACCTTTTCCAGAGCTCTG 928
QY 469 GGCCTTCCTCCGAGGAGCGTGAACCGG-----GGGTTGTGAAACGAGCCCTCAT 518
Db 929 GGGTTAGCCCAAGCAGCTGATGAGCCAGCGCCGAGAGATTCAACAGAGAGCAT 988
QY 519 GAGCTCGGAGGCAAGCTCTGCTGCTGCGGAAAGCGCCCGTTGGGGGCTGCCCCCTAG 578
Db 989 GAGGCTAAGGAGGCAAGTGTGTATCCCAAGAGCGCCAGTGTGACAGCTGCTGTGA 1048
QY 579 GGCCTTCGCGGAG 592
Db 1049 GAGCTGTGCGGAG 1062
```

```
RESULT 7
AAV35701
ID AAV35701 standard; cDNA; 1856 BP.
```

```
AC AAV35701;
```

```
XX 19-AUG-1998 (first entry)
```

```
XX CDNA encoding human MYH.
```

```
XX Human; MYH; hMYH; diagnosis; cancer; ss.
```

```
XX Homo sapiens.
```

```
XX Key Location/Qualifiers
```

```
XX FT CDS 172..1776
```

```
XX FT /tag= a "MYH"
```

```
XX FT /product= "MYH"
```

```
XX JPI0057076-A.
```

```
XX 03-MAR-1998.
```

```
XX 11-MAR-1997; 97JP-00099540.
```

```
XX 11-MAR-1996; 96US-0013132P.
```

```
XX (HUMA-) HUMAN GENOME SCI INC.
```

```
XX WPI; 1998-210407/19.
```

```
XX P-PSDB; AAW60243.
```

```
XX New isolated polypeptide - useful for, e.g. diagnosis of cancer.
```

```
XX Claim 6; Page 17-19; 23pp; Japanese.
```

```
XX The present sequence encodes human MYH (hMYH). The MYH polypeptide can be
```

```
XX used for diagnosing cancer, where the decreasing levels of the
```

```
XX polypeptide are measured in a sample from a patient
```

```
XX Sequence 1856 BP; 400 A; 545 C; 568 G; 343 T; 0 U; 0 Other;
```

```
Query Match 10.5%; Score 102.6; DB 2; Length 1856;
```

```
Best Local Similarity 54.4%; Pred. No. 4,4e-09;
```

```
Matches 291; Conservative 0; Mismatches 214; Indels 30; Gaps 3;
QY 88 GTCTGTCTCCGAGGCTCTTCTGACAGAGACCGGGGTGAGAGCCCTCCCTATTAC 147
Db 514 GTGTGGGTCTCAGAGGTATGATCTGACAGAGACCGAGGTTCACATGTATCACTATAT 573
QY 148 CGCCGCTTTTGGAGCGCTTCCACACCTGAAGCCCTGGGCGGCTTCCCTGGAGAG 207
Db 574 ACCGATGATGCAAGATGAGCTTACCTGAGAGACCTGGGCGAGCTTCCCTGGAGAG 633
QY 208 GTCTTTAGAGTCTGAGAGGAGGAGGCTTACCTGAGAGAGCTTCCCTGAGAGAG 267
Db 634 GTGATCAACTCTGGGCTGGGCTGAGTATCTGTGTGGCGGCGGCTGAGAGAGGA 693
QY 268 GCCGAGCGTGAAGAGCTTCCCCGAGCTT-----GCCGAGCTT 309
Db 694 GCTCGAAGAGTGTGAGAGAGCTAGGGGGCCACATGCCACGTACAGAGAGAGCCCTGAG 753
QY 310 CGGGGCTTCTGAGCTGGGCTTACACCGGGCGGCGGCTGAGAGAGCTTCCCTGAGAG 369
Db 754 CAGCTTCTGCTGGGCTGGGCTGAGTACAGAGCTGAGAGAGCTTCTGAGAGAGCTTGGC 813
QY 370 GAGCGGAGTGGGAGTGAAGAGAGCTCCGAGAGGCTCTGCGGCTTGGAGAG 427
Db 814 CAGGCAACGGGTGTGTGATGAGAGAGTACAGAGAGGCTGTGCGGCTGAGAGAGCTT 873
QY 428 GGGAAAGCCCAAGAGAGAGCTTTCGCTGCGGCTGAGAGAGCTTCCCTGAGAGAG 487
Db 874 GGTGTGTATCCACAGAGAGCTTGTTCAGAGAGCTTGGGGTCTAGCCACAGAGCTG 933
QY 488 TGG-----ACCGGGGAGTGAAGACAGAGCCCATGAGCTGAGAGAGAGAGCTC 537
Db 934 GTGAGCCAGAGCGGCGGAGAGATTTCACACAGAGAGCATGAGAGCTAGAGAGAGAG 993
QY 538 TGCCTGCGGAAAGCGCCCGCTTGGGAGAGCTGCGGCTTGGAGAGAGCTTGGCGAG 592
Db 994 TGTACCCAGAGCGGCGGAGCTGTGAGAGAGAGCTTGGAGAGAGCTGTGCGGAG 1048
```

```
RESULT 8
```

```
ADBE36257
```

```
ID ADBE36257 standard; cDNA; 1608 BP.
```

```
XX ADBE36257;
```

```
XX 29-JAN-2004 (first entry)
```

```
XX Human MYH (hMYH) A494G mutant cDNA.
```

```
XX gene; ss; mutant; human; screening method; hMYH; base excision repair;
```

```
XX BER; APC; familial adenomatous polyposis; FAP;
```

```
XX multiple colorectal adenoma; carcinoma; bowel cancer.
```

```
XX Homo sapiens.
```

```
XX Key Location/Qualifiers
```

```
XX FT CDS 1..1608
```

```
XX FT /tag= a
```

```
XX FT /product= "Human MYH mutant protein (Y165C)"
```

```
XX FT replace (494,a)
```

```
XX FT /tag= b
```

```
XX FT /note= "Base G in mutant replaces wild type base A"
```

```
XX WO2003014390-A2.
```

```
XX 20-FEB-2003.
```

```
XX 02-AUG-2002; 2002WO-GB003591.
```

```
XX 03-AUG-2001; 2001GB-00018995.
```

```
XX (UYWA-) UNIV WALES COLLEGE OF MEDICINE.
```

PI Sampson JR, Cheadle JP;  
 XX WPI: 2003-256601/25.  
 DR P-PSDB; ADE36260.  
 XX  
 PT Screening, diagnostic and therapeutic methods in individuals with  
 PT predisposition towards having a cancer, such as colon cancer, using base  
 PT excision repair pathway or hMYH gene.  
 XX  
 PS Claim 8; SEQ ID NO 1; 66pp; English.  
 XX  
 CC This invention relates to a novel screening method for identifying an  
 CC individual having a predisposition towards a cancer. Specifically, it  
 CC refers to obtaining a test sample, preferably comprising the hMYH gene  
 CC that occurs in the base excision repair (BER) pathway, and comparing this  
 CC nucleic acid molecule to the corresponding region of the wild type  
 CC sequence. This BER pathway gene, hMYH, acts to protect against G:C to T:A  
 CC transverse mutations in a cancer marker gene such as APC that is seen in  
 CC familial adenomatous polyposis (FAP). As such, mutations identified in  
 CC hMYH are associated with the onset multiple colorectal adenomas and  
 CC carcinoma. The present invention describes a screening method for  
 CC individuals that works to identify differences comprising any one of  
 CC G382D, Y165C, E466X or Y90X variations in hMYH, this signifies a cancer  
 CC predisposition, particularly for bowel cancer. This polymucleotide  
 CC sequence is the mutant A494G hMYH cDNA that encodes a Y165C mutation in  
 CC the encoded polypeptide of the invention.  
 CC  
 XX Sequence 1608 BP; 336 A; 477 C; 504 G; 290 T; 0 U; 1 Other;  
 SQ  
 Query Match 10.5%; Score 102; DB 9; Length 1608;  
 Best Local Similarity 53.6%; Pred. No. 5.7e-09;  
 Matches 297; Conservative 1; Mismatches 226; Indels 30; Gaps 3;  
 QY 69 GGAGAGAGACCCCTTACCGCGTCTGCTCCAGGCTCTTTCGACGACAGCCCGGTGA 128  
 DB 327 GGACAGGGGGGATATGCTGTGTGTCTCAGAGGTATGTCGACGACAGCCAGTTGC 386  
 QY 129 GCAGGCTTCCCTTATACCGCGCTTTTTCGAGCGCTTTCCACCTCGAAGCCCTTGC 188  
 DB 387 CACTGTATCAACTACTATACCGAGTGTATGACAGAGGAGGCTTACCTGAGGACTTGC 446  
 QY 189 CGCGGCTTCCCTGAGAGAGTCTTATAGGTCTGCGACGAGGGGGCTTACTCCCGCGGC 248  
 DB 447 CAGTCTTCTCTGAGAGAGGTATCACTTGGCTTGGCTTCTTATTTCTCTGCG 506  
 QY 249 GGAACACCTCCACCGCGCTGCGCGGAGAGGAGCTTCCCGAGCTTC----- 300  
 DB 507 CCGCGGCTGACAGAGGAGCTCGGAGGTGTAGAGAGGAGGCGGCACATGCCAG 566  
 QY 301 -----GCCGAGCTTGGGGGCTTCTGCTCTCGGCTTACACCGGCGCGCT 350  
 DB 567 TACAGCAGAGACCTTGCAGCAGCTCTGCTGCGGCGGCTTACACAGCTGGGGCAT 626  
 QY 351 GGCTTCATGCGCTTGGGGGAGCGGGTGGGCGGTGACGGGAACGTCGGAGGGTCT 410  
 DB 627 TGCCTTATGCTTGTGCGCAGCAACCGGTGTGTGATGCAACGAGACGGGTCT 686  
 QY 411 CTCGCGCTTCTGCGCC--GGGAAAGCCCAAGAGAGAGCTTTTGGCCCTGCGCAG 468  
 DB 687 GTGCGGTGTGCGAGCAGTGTGTGATCCAGCAGCAGCCTTGTTCGACGAGCTTG 746  
 QY 469 GGCCTCTCCCGAGGCGGTGACCGG-----GGGTGTGGAACAGGCGCTCAT 518  
 DB 747 GGGTGTAGCCAGAGCTGTGTGACCCAGCCCGGCAAGATTTCAACAGAGCGCAT 806  
 QY 519 GGAGCTGGGGGCGACGCTGTGCTGCGGAAAGCGCCCGTTCGGGGGCTGCGCCCTTACG 578  
 DB 807 GGAAGTAAAGGCGCAGTGTGTATCCCAAGCGCCCACTGTGACGCCAGTCCCTGTGA 866  
 QY 579 GGCCTTCTGCGGG 592  
 DB 867 GAGCTGTGCGGG 880

RESULT 9  
 ID AAF21610 standard; DNA; 349980 BP.  
 XX  
 AC AAF21610;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:111.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
 KW ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN MO20006791-A1.  
 XX  
 PD 09-NOV-2000.  
 XX  
 XX 08-MAR-2000; 2000WO-US005928.  
 PF  
 XX 30-APR-1999; 99US-0132068P.  
 PR 08-OCT-1999; 99WO-US023573.  
 PR 28-FEB-2000; 2000GB-00004695.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Pizsa M, Hickey E, Peterson J, Tettelin H, Venter JC, Scarlato V,  
 PI Maignani V, Galeotti C, Mora M, Ratti G, Scarlatti M, Scarlato V,  
 PI Rappuoli R, Frazee CM, Grandi G;  
 XX  
 DR WPI: 2000-647603/62.  
 XX  
 PT Neisseria meningitidis B full length genome sequence and open reading  
 PT frames are used to detect, treat and prevent Neisserial infections.  
 XX  
 PS Claim 7; Appendix A; 692pp; English.  
 XX  
 CC The present invention describes the full length genome of Neisseria  
 CC meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613  
 CC represent fragments of the NMB genomic sequence, as the sequences which  
 CC long to go in a record on its own it was split into 8 sequences which  
 CC overlap each other at the beginning and end of each sequence by 49980 bp  
 CC (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of  
 CC AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of  
 CC AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins  
 CC given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR  
 CC primers which are used in the exemplification of the present invention.  
 CC The NMB genome and fragments from it have antibacterial activity, and can  
 CC be used in vaccines and gene therapy. Neisseria nucleic acids, proteins  
 CC and/or antibodies which binds to the proteins can be used in compositions  
 CC for treating or preventing infection due to Neisserial bacteria or as a  
 CC diagnostic reagent for detecting the presence of Neisserial bacteria or  
 CC of antibodies raised to Neisserial bacteria. Computers, computer memory,  
 CC computer storage medium or computer databases can be used in a search to  
 CC identify open reading frames (ORFs) or coding sequences within the NMB  
 CC genome. The DNA sequences provide further opportunities to find antigenic  
 CC or immunogenic proteins which are more effective in vaccines than the  
 CC outer membrane proteins currently used  
 XX  
 SQ Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 U; 0 Other;  
 Query Match 9.8%; Score 95.6; DB 3; Length 349980;  
 Best Local Similarity 51.6%; Pred. No. 5.8e-08;  
 Matches 255; Conservative 0; Mismatches 224; Indels 15; Gaps 1;  
 QY 17 AAGCCTCTCTGCTGTGATCCGGAAGAAAGCGCCCGCTTCCCTGCGGGGAGAGG 76  
 DB 225615 AAGGCTCATCTCGCTGGCAAAAACACACCGGTGCGCACCACTCTGCGAGGTCAAAA 225674

QY 77 ACCCTTACCGGCTCTGGTCTCCGAGTCTTCTGACAGACGCCGGGTGAGACGGCC 136  
 Db 225675 ACCCTTATTTGGCTCTGGCTTCCGAATATGCTCCAGAAACGCAAGTCGCACCGTGT 225734  
 QY 137 TCCCTTATTTACCGGCTTTCTGAGCCCTTCCACCTGAAAGCCCTTGCGCGGCTT 196  
 Db 225735 TGGACTACTATCCGGCTTCTTGAATAATTCGCCACGCTTGAACGCTTGCGCGGCGC 225794  
 QY 197 CCTTGAAGAGGTCTTGGGTCTGGCAGAGGGCGGGCTACTACGGCGCGGCGAAACCC 256  
 Db 225795 CGCAAGACGAAGT 225854  
 QY 257 TCCACCGCTGGCGCCGAAAGCTGAGAGAC-----TTCGCCGAGCTTGC 301  
 Db 225855 TGCACAAAGCCCGCAACAAAGTGTGAGCAATTCGGCGGCAAGTTTCGTCGAGCGCA 225914  
 QY 302 CGGAGCTTGGGGGCTTCTGTCTGCGGCTTACACCGCGCGCGCGGCTCTCCATCG 361  
 Db 225915 AAGACTTGAACCCCTCTGCGGCTGAGCAAGACCGCGCGCAATTTGCGCGCTTCT 225974  
 QY 362 CCTTGGGAGAGGGGGTGGCGGGGTGAGCGGAACTGCCGAGGCTCTCTCCGCTCT 421  
 Db 225975 CCTTCAACCGCGCGCAACATTTTGAAGCAAGCTCAAGCGTACTCTGCGCGTGT 226034  
 QY 422 TCGCCCGGAGAAAGCCCAAGAGAGAGAGCTTTTGCCTCGCCAGGCGCTTCTCCCG 481  
 Db 226035 TCGCCCGGAGCAATCCGAGAGCAAAATTTGAAATCTGCTTGAACATTTGCCG 226094  
 QY 482 AGGCGCTGAGACCG 495  
 Db 226095 AAGCCTGCTGCCG 226108

RESULT 10  
 AAA81490\_13  
 Continuation (14 of 15) of AAA81490 from base 1300001 (N. meningitidis B full length gen  
 WP Sequence split into 15 fragments LOCUS AAA81490 Accession Aa81490

Fragment Name	Begin	End
WP AAA81490_00	1	110000
WP AAA81490_01	100001	210000
WP AAA81490_02	200001	310000
WP AAA81490_03	300001	410000
WP AAA81490_04	400001	510000
WP AAA81490_05	500001	610000
WP AAA81490_06	600001	710000
WP AAA81490_07	700001	810000
WP AAA81490_08	800001	910000
WP AAA81490_09	900001	1010000
WP AAA81490_10	1000001	1100000
WP AAA81490_11	1100001	1210000
WP AAA81490_12	1200001	1310000
WP AAA81490_13	1300001	1410000
WP AAA81490_14	1400001	1437668

Query Match 9.8%; Score 95.4; DB 3; Length 110000;  
 Best Local Similarity 51.7%; Pred. No. 6.6e-08;  
 Matches 253; Conservative 0; Mismatches 221; Indels 15; Gaps 1;

QY 22 CTCTCTCGCTGTGATCCGGAAGAGCCGCGCTCTCTCTGCGGGGAGAAAGACCT 81  
 Db 96008 CTCTCTCGCTGTGCAAAAACAAACAGGTGCGCACACCTCTCTGAGGTCAAAAACCT 96067  
 QY 82 TACCGGCTCTGTCTCCGAGTCTTCTGAGAGACCCGGGTGAGAGCGCCCTCCC 141  
 Db 96068 TATGCGCTGTGCTTCCGAATATCTCTCAGCAAAAGCAAGTGCACCGTGTGAC 96127  
 QY 142 TATTAACGCGGCTTCTGAGAGCTTCCACCTGAAAGGCTCGCGCGGCTTCCCTG 201  
 Db 96128 TACTATTCGCGCTTCTTGAATAATTCGCCACGCTTCAAGAGCTTGCCTGCGCGCA 96187  
 QY 202 GAAGAGTCTTAGAGTCTGACAGAGGCGGCTACTACCGCGGGCGAAACACTTCAAC 261  
 Db 96188 GACGAGT 96247

QY 262 CGCTGCGCCGAAGCTGAGAGAC-----TTCGCCGAGCTTGCGCGAG 306  
 Db 96248 AAAGCCCGCAACAAAGTGTGAGCAATTCGCGGCAAGTTTCCGTCGAGCGCAAGAC 96307  
 QY 307 CTTCGGGGGCTTCTGTGCTCTGAGCTTACACCGCGCGCGGTGAGCTTCATGCGCTTC 366  
 Db 96308 TTGAAACCTCTGCGCGCTGAGCAGAGACCGCGCGCCATTTGCGCTTCTCTTC 96367  
 QY 367 GGGGAGCGGTGCGCGGTGAGACGAGACGTCGAGAGGTCTCTTCCGCTTTCGCC 426  
 Db 96368 AACCGCGCAACATTTTGAACGCGCAACGCTACTCTGCGCGGTTCGCC 96427  
 QY 427 CGGAAAGCCCAAGAGAGAGAGCTTTTGCGCTGCGCCAGGAGCTCTCCCGAGGCG 486  
 Db 96428 CGCAGCGCAATCCGAGCAAAATTTGAAATCTGCTTGAACATTTGCCGAAAGC 96487  
 QY 487 GTGAGCCCG 495  
 Db 96488 CTGCTGCCG 96496

RESULT 11  
 ABK74292  
 ID ABK74292 standard; DNA; 528 BP.  
 XX AC ABK74292;  
 XX 13-AUG-2002 (first entry)  
 DT Bacillus licheniformis genomic sequence tag (GST) #1583.  
 DE Bacillus licheniformis genomic sequence tag (GST) #1583.  
 KW Differential gene expression; genomic sequenced tag; GST;  
 KW altered culture condition; environmental stress;  
 KW physiological provocation; ds.  
 OS Bacillus licheniformis.  
 XX  
 XX PN WO200229113-A2.  
 XX PD 11-APR-2002.  
 XX PF 05-OCT-2001; 2001WO-US031437.  
 PR 06-OCT-2000; 2000US-00680598.  
 PR 27-MAR-2001; 2001US-0279526P.  
 XX (NOVO) NOVOZYMES BIOTECH INC.  
 XX (NOVO) NOVOZYMES AS.  
 XX Berke R. Clausen IG;  
 XX WPI; 2002-416684/44.  
 XX  
 PT Monitoring differential expression of several genes in first Bacillus  
 PT cell relative to expression of same genes in one or more second Bacillus  
 PT cells, by using substrate containing Bacillus genomic sequenced tag  
 PT array.  
 XX  
 XX Claim 4; SEQ ID NO 1583; 2000pp; English.  
 XX  
 CC The invention describes a method of monitoring differential expression of  
 CC genes in a first Bacillus cell relative to expression of the genes in  
 CC other Bacillus cells, comprising hybridizing labelled nucleic acid probes  
 CC isolated from Bacillus cells to a substrate containing array of Bacillus  
 CC genomic sequenced tags (GST), examining the array, and determining  
 CC relative gene expression by an observed hybridisation reporter signal of  
 CC a spot in the array. The method is useful for measuring the expression of  
 CC genes in a first Bacillus cell relative to expression of the same genes  
 CC in one or more second Bacillus cells. The method is useful for monitoring  
 CC global expression of several genes from a Bacillus cell, discovering new  
 CC genes, identifying possible functions of unknown open reading frames and  
 CC monitoring gene copy number variation and stability. Monitoring changes

CC in expression of genes may be used to provide a representation of the way  
CC in which *Bacillus* cells adapt to changes in culture conditions. Extensive follow  
CC environmental stress or other physiological provocation. Extensive follow  
CC -up characterisation is unnecessary, when one spot on an array equals one  
CC gene or one open reading frame, since sequence information is available.  
CC This sequence represents a genomic sequence tag (GST) used in the method  
CC of the invention. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPRO at [http://wipro.int/pub/published\\_pct\\_sequences](http://wipro.int/pub/published_pct_sequences)

XX Sequence 528 BP; 160 A; 111 C; 140 G; 116 T; 0 U; 1 Other;

Query Match 9.5%; Score 92.8; DB 6; Length 528;

Best Local Similarity 53.4%; Pred. No. 2.3e-07;

Matches 227; Conservative 0; Mismatches 183; Indels 15; Gaps 1;

```
QY 17 AAGCCCTCTGCGCTGTACCGGGAACGCGCCGCCCTCCCTGCGGGGGGAGAG 76
   |||
DB 43 AAGATTATTTCTTGTATGAGCAAGAAAACGGGATCTGCTTGGCGCTCAGACAG 102
   |||
QY 77 ACCCTTACCGGCTCTGCTCCGAGCTCTTTCAGACAGACCCGGGTGAGCAGGCC 136
   |||
DB 103 ACCCTTATAGGTGTGGGTGTGGAAGTATGCTTCAAGACGAGGTGACAGGCTGA 162
   |||
QY 137 TCCCTATTACCGCGCTTTCGAGCGCTTCCACCGCTGAAGGCCCTGCGCGGCTT 196
   |||
DB 163 TTCTTACTTCAACAACCTTATCGAAAGTTTCGACCGTGAAGCGCTCGTGAAGCG 222
   |||
QY 197 CCTTGAAGAGGTCTTGAAGGTCTGCGAGGCGGCGGCTACTACCGGCGGCGAACC 256
   |||
DB 223 ACGAAGAAAAGTGTCTTAAAGCTGGAAGACGTGGATATCTCAAGGGTTGGAACT 282
   |||
QY 257 TCACACCG-----CTGCCCGAAGCTGGAAGGCTTCCCGAGCTTCG 301
   |||
DB 283 TGCAGAGCGCTGTACAGGAAGTTCAAGACATACGAGGTGTGCTCCCTTCAGAAAG 342
   |||
QY 302 CCGAGCTTCGCGGGCTTCTGTCTCGGCGCTTACACCGCGCGGCTGAGCTTCATCG 361
   |||
DB 343 AGGAATTCGCAAGCTTAAAGGCGTGGCCCTTATACAAAGGCGGCTGAGCATCG 402
   |||
QY 362 CTTTCGGGAGCGGCTGCGGCGGTGAGCGGAAAGCTCCGAGGCTCTCCCGCTCT 421
   |||
DB 403 CCTTCAATACAGCGGCTCCCGCGGTAGACGGAATGTATCGGCTCATGTACGAGATTC 462
   |||
QY 422 TCGCC 426
   |||
DB 463 TTTC 467
```

RESULT 12  
ABZ40024  
ID ABZ40024 standard; DNA; 1047 BP.

XX ABZ40024;

XX 07-MAR-2003 (first entry)

DE N. gonorrhoeae nucleotide sequence SEQ ID 4637.

XX Antibacterial; infection; vaccine; gene therapy; gene; ds.

XX *Neisseria gonorrhoeae*.

XX MO200279243-A2.

XX 10-OCT-2002.

XX 12-FEB-2002; 2002WO-IB002069.

XX 12-FEB-2001; 2001GB-00003424.

XX (CHIR-) CHIRON SPA.

XX PA

PI Fontana MR, Pizza M, Maignani V, Monaci E;  
XX WPI; 2003-058415/05.  
DR P-PSDB; ABP79054.

PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a  
PT medicament for treating or preventing *N. gonorrhoeae* infection.

XX Disclosure; Page 528; 815pp; English.

XX The present invention relates to proteins from *Neisseria gonorrhoeae*.

CC Also disclosed are the nucleic acid molecules encoding the proteins and  
CC antibodies that specifically bind to the proteins. The composition  
CC comprising the protein, nucleic acid or antibody is useful for the  
CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*  
CC infection, this may be in the form of a vaccine or gene therapy.

CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid  
CC molecules of the invention

XX Sequence 1047 BP; 259 A; 306 C; 262 G; 220 T; 0 U; 0 Other;

XX Query Match 9.1%; Score 89.2; DB 7; Length 1047;

Best Local Similarity 50.8%; Pred. No. 9.5e-07;

Matches 251; Conservative 0; Mismatches 228; Indels 15; Gaps 1;

```
QY 17 AAGCCCTCTGCGCTGTACCGGGAACGCGCCGCCCTCCCTGCGGGGGGAGAG 76
   |||
DB 35 AAGGCTCATCGCTGGAAGAAAACACAGGTGTGCGACACCTCTTGGCAGGTCAAA 94
   |||
QY 77 ACCCTTACCGGCTCTGCTCCGAGGTCTTTCGACAGACCCGGGTGAGCAGGCC 136
   |||
DB 95 ACCCTTATGGGTGTGCTTTCGAAATCATGCTCAGCAACGAGGTCCCGCGGTGT 154
   |||
QY 137 TCCCTATTACCGCGCTTTCGAGCGCTTCCACCGCTGAAGGCTTGCCTCGCGCTT 196
   |||
DB 155 TGAATCTATTCGCGCTTCTTGGAAAATTCGACCGCTTCAAGCGCTTCCCGCGCG 214
   |||
QY 197 CCTTGAAGAGGTCTTGAAGGTCTGCGAGGCGGCGGCTACTACCGGCGGCGGAACC 256
   |||
DB 215 CGAAGAGAGAGTGTGTGTGTGGGCGGCTTGTATTAACGCGCGCGCGCAATC 274
   |||
QY 257 TCACAC-----CGCTTGGCGCGAAGCTGGAAGGCTTCCCGAGCTTCG 301
   |||
DB 275 TGCACAAAGCGCGCAACAAATCTGCGACAAATTCGGGGTATCTTCATCGAGGCCA 334
   |||
QY 302 CCGAGCTTCGCGGGCTTCTGTCTCGGCGCTTACACCGCGCGGCTGAGCTTCATCG 361
   |||
DB 335 AAGACTTGAAGAACGCTCGCGCGGTAGGACAGACACGCGCGCATTTCTGCTTTG 394
   |||
QY 362 CTTTCGGGAGCGGCTGCGGCGGTGAGACGGGAACGTCCGAGGCTCTTCCCGCTCT 421
   |||
DB 395 CTTTCAACCGACGAGAAACCATCTTGAACGCGCAACGCGGTCTTCTGCGGTGTT 454
   |||
QY 422 TCGCCCGGGAAGCGCCAGAGAGAGCTTTTCGCGCTCGCCAGGCGCTCTCCCGCG 481
   |||
DB 455 TTGCGCAAGACGCGCAATCCGACATTAATAAATTTGAAAACTCGCTCTGACACTTGGCG 514
   |||
QY 482 AGGCGGTGACCCG 495
   |||
DB 515 AAGCGCTGATGCGG 528
```

RESULT 13  
AAS63242  
ID AAS63242 standard; cDNA; 2293 BP.

XX AAS63242;

XX 29-JAN-2002 (first entry)

XX *Escherichia coli* (strain K-12) A/G-specific adenine glycosylase MutY DNA.

XX DNA mutation-binding protein; nuclease; DNA mismatch; cancer; PCR primer;

XX KM

KM DNA damage; human xeroderma pigmentosum complementation group; XPE, XPA;  
 KM XPC; XPE; BECC4; human Muts homologue 2, hMSH2; Muts; Nuc; MutY; Fpg; ss;  
 KM Fapy-DNA glycosylase; uracil DNA glycosylase; ung; TDG; xtha gene; Uvr A;  
 KM A/G-specific adenine glycosylase; synthetic T4 endonuclease V; T4 endo V;  
 KM thymine DNA-glycosylase; Uvr B; Uvr C; nh gene; nfo gene; exonuclease;  
 KM endonuclease.  
 OS *Escherichia coli*.  
 XX  
 XX MO200173079-A2.  
 PN  
 XX  
 XX 04-OCT-2001.  
 PD  
 XX  
 XX 26-MAR-2001; 2001WO-US009700.  
 XX  
 XX 28-MAR-2000; 2000US-0192764P.  
 PR  
 XX 29-AUG-2000; 2000US-00650855.  
 XX  
 XX (BEGC) UNITV CALIFORNIA.  
 PA  
 XX  
 XX Mc Cutchen-Maloney SL;  
 PI  
 XX  
 XX WPI: 2001-656920/75.  
 DR  
 XX P-PsDB; AAU69755.  
 XX  
 XX Recombinant chimeric protein, useful for detecting and quantifying DNA  
 PT mutations, e.g. in disease diagnosis, comprises mutation-binding protein  
 and nuclease.  
 PS  
 XX  
 XX Claim 58; Page 99-100; 128pp; English.  
 CC  
 CC Sequences AAS63224-AAS63305 represent cDNA encoding and PCR primers used  
 CC to amplify cDNA encoding proteins which can be used in the synthesis of  
 CC chimeric proteins comprising a DNA mutation-binding protein, a linker and  
 CC a nuclease, by recombinant technology. The chimeric proteins are useful  
 CC for detection, quantification and mapping of DNA sequence variations  
 CC including mutations, for example, caused by damage and mismatches. The  
 CC proteins are able to bind to the site of the DNA mutation and cut it out  
 CC of the molecule. This is useful for early diagnosis of cancer and other  
 CC diseases. The proteins used in the invention include human XPF (or  
 CC ERCC4), human xeroderma pigmentosum complementation groups A, C and E  
 CC (XPA, XPC and XPE), human Muts homologue 2 (hMSH2), *Serratia marcescens*  
 CC nuclease (Nuc), *Thermus thermophilus* Muts, *Escherichia coli* Fapy-DNA  
 CC glycosylase (Fpg), uracil DNA glycosylase (ung), A/G-specific adenine  
 CC glycosylase (MutY), synthetic T4 endonuclease V (T4 endo V), thymine DNA-  
 CC glycosylase (Tdg), *E. coli* Uvr A, B and C, and *E. coli* endonucleases and  
 CC exonucleases  
 CC  
 XX  
 XX Sequence 2293 BP; 533 A; 620 C; 593 G; 547 T; 0 U; 0 Other;  
 SQ  
 Query Match 8.9%; Score 87.2; DB 4; Length 2293;  
 Best Local Similarity 54.1%; Pred. No. 2e-06;  
 Matches 210; Conservative 0; Mismatches 163; Indels 15; Gaps 1;  
 QY 53 CCTCTCCCTGGCGGGGGAAGAGACCTTACCGGCTCCTGCTCCGAGGCTCTTTCG 112  
 DB 1161 CTCTGCTCCCTGGCAATTGACAAGACGCTTACAAAGTAGGTCTCACAAGTAGTTC 1220  
 QY 113 AGCAGACCGGGGGGAGCAGCGCTTCTTATTCAGCGGCTTTTGAGAGCGCTTCCCA 172  
 DB 1221 AACAACTAGAGTGAGACCGTTATCCCTATTGGAACGCTTTATGCGCGGCTTCCGA 1280  
 QY 173 CCTGAAAGCCCTGGCGCGGCTTCCGGAAGAGTCTTAAAGGTCTGAGAGGGGGGG 232  
 DB 1281 CGGTGACCGATCTCGGCATATGCGCGCTCGACGAAGTCTTCACCTGTGAGCCGGCTTG 1340  
 QY 233 GTTACTACCGGGGGGGAACCTTCAACCGCTTGCCCGGAAGCGTGGAG----- 282  
 DB 1341 GTTATTTAGCCCGCGCGCAATCTGCAATTAAGCGGCAACAAGTGGCGACTTTACAG 1400  
 QY 283 -----GAGCTTCCCGGAGCTTCCGCGAGCTTCCGGGGGCTTCGAGTCTCGGGCTTACA 337  
 DB 1401 GCGGTAAATTCGCGAATCTTTAGAGAGATTGACGACCTGCGGGGCGTGGGCTTCCA 1460

QY 338 CCGGCGGCGGCGGCTTCATTCGCTTGGGAGACCGGATGCGCGGTGACCGGAACG 397  
 DB 1461 CCGCAAGCGCGATTTCTTCTGCTTCTGCTTGGTAAACACTTTCGATTCGACGGAACG 1520  
 QY 398 TCCGAGAGGTCTCTCCGCGCTTTCGC 425  
 DB 1521 TCMAACGCTGCTGGCGCGCTGTATGC 1548  
 RESULT 14  
 ABZ79956  
 ID ABZ79956 standard; DNA; 912 BP.  
 XX  
 XX AC ABZ79956;  
 XX  
 XX DT 19-MAY-2003 (first entry)  
 XX  
 XX DE Mycobacterium tuberculosis multy ORF sequence SEQ ID NO.31.  
 XX  
 XX KM Mycobacterium tuberculosis; mutR2; alkA; cgt; Rv3908; mutY; Rv3909;  
 KM detection; multidrug resistance; multiple drug resistance; MDR;  
 XX infection; gene; ds.  
 XX  
 XX OS Mycobacterium tuberculosis.  
 XX  
 XX PN WO2003016562-A2.  
 XX  
 XX PD 27-FEB-2003.  
 XX  
 XX PF 14-AUG-2002; 2002WO-EP009679.  
 XX  
 XX PR 14-AUG-2001; 2001US-0311824P.  
 PR 21-AUG-2001; 2001US-0313523P.  
 XX  
 XX (INSP) INST PASTEUR.  
 XX  
 XX PI Gicquel B;  
 DR WPI: 2003-256711/25.  
 XX  
 XX PT Predicting the epidemic character of a Mycobacterium tuberculosis isolate  
 PT and/or the acquisition of multiple drug resistance (MDR) by the isolate  
 XX by detecting an alteration in the DNA repair system of the isolate.  
 PS Claim 45; Fig 6B; 83pp; English.  
 XX  
 XX CC The present invention describes a method for predicting the epidemic  
 CC character of a Mycobacterium tuberculosis isolate and/or a selective  
 CC advantage to be maintained in the host and/or the acquisition of multiple  
 CC drug resistance (MDR) by the isolate comprising detecting an alteration  
 CC in the DNA repair system of the isolate. Also described: (1) detecting a  
 CC polynucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (2) a  
 CC an *Escherichia coli* strain containing the plasmid pMYC2501; and (5)  
 CC detecting in a patient infected by Mycobacterium tuberculosis a higher  
 CC risk of being unable to eliminate the bacillus or of developing MDR  
 CC tuberculosis. The method is useful for predicting the epidemic character  
 CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to  
 CC be maintained in the host and/or the acquisition of MDR by the isolate.  
 CC The present sequence represents a M. tuberculosis multy open reading frame  
 CC (ORF) sequence, which is used in the exemplification of the present  
 CC invention  
 CC  
 XX  
 XX Sequence 912 BP; 124 A; 311 C; 326 G; 151 T; 0 U; 0 Other;  
 SQ  
 Query Match 8.9%; Score 86.6; DB 7; Length 912;  
 Best Local Similarity 50.2%; Pred. No. 2.7e-06;  
 Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;  
 QY 22 CTCTGCTGCTGATACCGGAAAGACCGCGCCCTCCCGCTGGGGGGG---GAGAAAGAC 78  
 DB 61 CTTCGCTTGATACGACATCGACCGGAGTCTGCTGGGAGAGCCCGGTGTACG 120

```
OY 79 CTTACCGCGCTGCTCCGAGGTCCTTCTGACAGACCCGGGTGAGAGCCCTC 138
DB 121 CCGTGGCAGATCTGTCGACGAGTTCACTGTCGACGACGCCGCCGCCGGGTGCTG 180
OY 139 CCTATTACCGCCGCTTTCTGAGAGCCTTTCACCTGAAAGCCCTGCGCGCTTCC 198
DB 181 GCGATCTGGCCGGAATGGGTGGCGGTGGCCCAAGCCGCTGAGCCACCGCCAGC 240
OY 199 CTGGAAGAGTCTTAGGCTCTGCGAGGGGGGCGCTACTACCGCGCGGAAACCTC 258
DB 241 ACCCGCATGTATTACGCGCTGGGGCAAGCTGGGCTATCCAGCGCAGCAAGCGCTTA 300
OY 259 CACCGCTGGCCGGAAGGCTG-----GAGAGCTTCCCGCAGCTTCCGC 303
DB 301 CACGATGCGCACCGTATCGCCCGACCAACAATGAGTGTGCTCCGAGATATTCAG 360
OY 304 GAGCTTGGGGGCTTCTGTGCTCGAGGCTTACACCGCGCGCGGCTCCATCGCC 363
DB 361 ATCTGTGTCACCTTGCCTGGGCGCTCGGAGCTACACCGCGCGCGGCTGTTGCT 420
OY 364 TTGCGGAGCGGGTGGCGGCGGTGACGGAACGTCCGAGGCTTCTCCGCTTT- 422
DB 421 TACCGCAGCGGGTGGCGGTGTGACACCAATGTGCGCGCTGTGCTGCCGCGCTT 480
OY 423 -----CGCCCGGAAAGCCCAAGAGAGCTTTGCGCTTCCGCTCGCCAGGCG 471
DB 481 CACGCGCGCGCGGACCGCGGTGCGCATCGGTGCGCGCGACCAAGCGAGCTCTTGCG 540
OY 472 CTCCTCCCGAGGCGGTGACCCGGGGGTGGAACACAGGCGCTCATGAGCTCGGGGC 531
DB 541 CTGTGCGCGACCGCGAGACGGCGCTGAATTTTCGTGCGCTGATGAGTTGGGTGG 600
OY 532 ACGGTCTGCTGCGGAACCGCGCGCTTTCGAGGCGCTGCGCGCTTTCGCGG 591
DB 601 ACGGTGTGACCGCGCGACCGCGGTGCGGTATGCGCGTGGATGGTGGCATGG 660
OY 592 GGGAGAGGCGCGCGCGCTACCCCGCGCGCGAGAGCGCGCGGCAAGAGAGCGC 651
DB 661 CGGATGCGCGGTATTCGCGCGCTCGGACGCTC---GCCGCGCGGGGCAAGCGCTTAC 717
OY 652 CTCGTGCGCTGCTCTCTCTCTCGGGCG 678
DB 718 GGAACCGACCGCAAGTCCGCGACGG 744

RESULT 15
AB279955
ID AB279955 standard; DNA; 1312 BP.
XX
AC AB279955;
XX
DT 19-MAY-2003 (first entry)
XX
DE Mycobacterium tuberculosis muty nucleotide sequence SEQ ID NO:21.
XX
KM Mycobacterium tuberculosis; mutY2; alkA; ogc; RV3908; mutY; RV3909;
KM detection; multidrug resistance; multiple drug resistance; MDR;
XX infection; gene; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN MO2003015662-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002MO-EP009679.
XX
PR 14-AUG-2001; 2001US-0311824P.
XX 21-AUG-2001; 2001US-0313523P.
XX
PA (INSP ) INST PASTEUR.
XX
```

```
PI Gilequel B;
XX
DR WPI: 2003-256711/25.
XX
PT Predicting the epidemic character of a Mycobacterium tuberculosis isolate
PT and/or the acquisition of multiple drug resistance (MDR) by the isolate
PT by detecting an alteration in the DNA repair system of the isolate.
XX
PS Disclosure; Fig 6A; 83pp; English.
XX
CC The present invention describes a method for predicting the epidemic
CC character of a Mycobacterium tuberculosis isolate and/or a selective
CC advantage to be maintained in the host and/or the acquisition of multiple
CC drug resistance (MDR) by the isolate comprising detecting an alteration
CC in the DNA repair system of the isolate. Also described: (1) detecting a
CC Mycobacterium tuberculosis strain with a MDR phenotype; (2) a
CC polynucleotide; (3) a kit for detecting Mycobacterium tuberculosis; (4)
CC an Escherichia coli strain containing the plasmid pMYC2501; and (5)
CC detecting in a patient infected by Mycobacterium tuberculosis a higher
CC risk of being unable to eliminate the bacillus or of developing MDR
CC tuberculosis. The method is useful for predicting the epidemic character
CC of a Mycobacterium tuberculosis isolate and/or a selective advantage to
CC be maintained in the host and/or the acquisition of MDR by the isolate.
CC The present sequence represents a M. tuberculosis muty nucleotide
CC sequence, which is used in the exemplification of the present invention
XX
Sequence 1312 BP; 173 A; 471 C; 457 G; 211 T; 0 U; 0 Other;
```

```
Query Match 8.9%; Score 86.6; DB 7; Length 1312;
Best Local Similarity 50.2%; Pred. No. 2.7e-06;
Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;
```

```
OY 22 CTCCTGCTGTGATACGGGAAAGCCCGCCCTCTCCCTGGCGGGG--GAGAAGAC 78
DB 261 CTTCTCGCTGTGATACGATCGACCGGATCTGCTTGGCGAGACCGCGTGTACG 320
OY 79 CTTACCGCGTCTGTGCTCGAGTCTTCTGAGAGAACCGCGGTGAGACGCTTC 138
DB 321 CCGTGGAGATCTGTGACGAGTTCATGTGAGAGACCGCGCGCGCGGTGTG 380
OY 139 CCTATTACCGCGCTTCTGAGAGCTTTCACACCTGAAAGCGCTTCCGCGCTTCC 198
DB 381 GCGATCTGGCGGACGTGGGTGGCGGTGCGCGCTGCGCGCGCGCGCGCGCGCG 440
OY 199 CTGGAAGAGTCTTAGGCTGTGCGAGGGCGGGCTACTACCGCGCGCGGAACACTTC 258
DB 441 ACGCGCATGTGTTACCGCGCTGGGGCAAGTGGCTATCCAGACGACCAAGCTTA 500
OY 259 CACCGCTGGCCGGAACGCTG-----GAGAGCTTCCCGCAGCTTGGCC 303
DB 501 CACGATGCGCGCACCGTATCGCCCGACCAACAATGACGTGTGCGCGGAGATATGAG 560
OY 304 GAGCTTGGGGGCTTCTGAGTCTGAGGCTTACACCGCGCGCGGTGAGCGCTTCATGCGC 363
DB 561 ATCTGTGTCACCTTGGCGGGGTGAGAGCTACACCGCGCGCGGTGAGCTGTTCCT 620
OY 364 TTGCGGAGGCGGTGCGCGGTGAGACGGAACCTTCGAGAGGCTCTCTCCGCTTT- 422
DB 621 TACCGCAGCGGGGTGGCGGTGTGACCAATGTGCGGGCGGTGTGGCCGCGCGCTT 680
OY 423 -----CGCCCGGAAAGCCCAAGAGAGACTTTTGGCTTGGCCAGGCG 471
DB 681 CACGCGCGCGCGACCGCGGTGCGCATCGGTGCGCGCGACCAAGCGCTTGTGGCG 740
OY 472 CTCCTCCCGAGGCGGTGACCCGGGGGTGGAACACGAGCGCTCATGAGCTCGGGGC 531
DB 741 CTGTGCGCGACCGCGAGACGGCGCTGAATTTTCGTGCGCTGATGAGTGGGTGG 800
OY 532 ACGGTCTGCTGCGGAACCGCGCGGTGCGGGCGCTGCGCTTGAAGGCGCTTCTGCGG 591
DB 801 ACGGTGTGACCGCGCGACCAACCGCGGTGCGGTATGCGCGCTGTGAGCTGTGCGCATGG 860
OY 592 GGGAGAGGCGCGCGCGCTTACCGCGCGCCCAAGAGAGCGCGCGGAGAGAGCGCG 651
```

Db	861	CGGCA	TGCCCGGTTATCCGCCGTCGAGCGGTC	917
Qy	652	CTCGT	CGCCCTCGTCCTCTCGGCGG	678
Db	918	GGACCG	ACCGCAAGTCCCGGACCG	944

Search completed: March 4, 2004, 04:19:18  
Job time : 366 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 4, 2004, 03:56:41 ; Search time 2811 Seconds

(without alignments)  
15033.603 Million cell updates/sec

Title: US-09-938-901-1

Perfect score: 975  
Sequence: 1 gtctgagcgccgcgcgaagc.....tagtccccctccgcgaagca 975Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

```
GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_un.*
13: gb_vl.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rnd.*
36: em_hcg_mam.*
37: em_hcg_vtc.*
38: em_sy.*
39: em_hgo_hum.*
40: em_hgo_mus.*
41: em_hgo_other.*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	975	100.0	975	6	BD174408
2	176.4	18.1	1089	1	BD174408 DNA resto
3	176.4	18.1	11378	1	AF377342 Deinococc
4	163.2	16.7	8249	1	AF543480 Deinococc
5	157.8	16.2	303642	1	AF543480 Uncultured
6	144.2	14.8	190050	1	AE016923 Chromobac
7	135	13.8	257701	1	AE640422 Ralstonia
8	135	13.8	346259	1	EX640435 Bordetell
9	135	13.8	346287	1	EX640435 Bordetell
10	132.2	13.6	293050	1	SC0939116 Mesorhizo
11	130.2	13.4	346547	1	AP003012 Streptomy
12	129.4	13.3	1106	1	AF121797 Caulobact
13	127.8	13.1	11283	1	AE005710 Rhodospir
14	124.4	12.7	345012	1	EX572607 Halobacte
15	123.8	12.7	11034	1	AE005065 Pseudomon
16	123.8	12.7	11786	1	AE004927 Bacteroid
17	118	12.1	300531	1	AE016932 Bacteroid
18	116.6	12.0	10923	1	AE014358 Brucella
19	116.4	11.9	299700	1	AP005944 Brucella
20	115	11.8	9829	1	AE009581 Brucella
21	114.8	11.8	349970	1	EX571659 Sequence
22	113	11.6	1551	6	AR388059 Sequence
23	112.4	11.5	299800	1	AP005040 Streptomy
24	111.6	11.4	303438	1	AE017212 Geobacter
25	110.4	11.3	310613	1	AE016875 Pseudomon
26	109	11.2	299950	1	AP005372 Thermosyn
27	108.8	11.2	286550	1	SMES91785 Simorhizo
28	107.8	11.1	300511	1	AE016775 Pseudomon
29	107.2	11.0	302988	1	AE017176 Porphyrom
30	105.8	10.9	299700	1	AP006571 Geobacter
31	105.4	10.8	13816	1	AE019335 Yersinia
32	105.4	10.8	208050	1	AJ414145 Yersinia
33	104	10.7	1688	9	AB032925 Homo sapi
34	104	10.7	1692	9	AB032929 Homo sapi
35	104	10.7	1721	9	AB032924 Homo sapi
36	104	10.7	1756	9	AB032928 Homo sapi
37	104	10.7	1757	9	AB032923 Homo sapi
38	104	10.7	1759	9	AB032927 Homo sapi
39	104	10.7	1821	9	AB032922 Homo sapi
40	104	10.7	1824	9	AB032921 Homo sapi
41	104	10.7	1838	9	AB032926 Homo sapi
42	104	10.7	1844	9	BC003178 Homo sapi
43	104	10.7	1854	9	AB032920 Homo sapi
44	104	10.7	1869	9	U63328 Human mult
45	102.8	10.5	348764	1	BX569689 Synechoco

## ALIGNMENTS

```
RESULT 1
LOCUS      BD174408
DEFINITION DNA restoration enzyme gene.
ACCESSION  BD174408
VERSION    BD174408.1 GI:29120096
KEYWORDS   JP 2002247985-A/1.
SOURCE     Thermus thermophilus
ORGANISM   Thermus thermophilus
            Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
            Thermus.
REFERENCE  1 (bases 1 to 975)
AUTHORS   Kuramitsu,N. and Yokoyama,S.
TITLE     DNA restoration enzyme gene
JOURNAL   Patent: JP 2002247985-A 1 03-SEP-2002;
```

COMMENT THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH  
OS Thermus thermophilus  
PN JP 2002247985-A/1  
PD 03-SEP-2002  
PF 23-FEB-2001 JP 2001047762  
PI NARUKI KIRAMITSU, SHIGEKU YOKOYAMA  
PC C12N15/09, C07K14/195, C12N1/15, C12N1/19, C12N1/21, C12N5/10 PC  
, C12N9/14, C12N9/16,  
PC C12N9/78, C12P21/02, C12N15/00, C12N5/00  
CC DNA restoration enzyme gene  
FH Key Location/Qualifiers  
FT CDS (1)..(975).

FEATURES

source

ORIGIN

Query Match 100.0%; Score 975; DB 6; Length 975;  
Best Local Similarity 100.0%; Pred. No. 8,4e-121;  
Matches 975; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTGGAGGCGCTGGGGGAAAGCCCTCCGCTGGTACCGGGAAAGCGCCGCCCTCCCGCC 60  
1 GTGGAGGCGCTGGGGGAAAGCCCTCCGCTGGTACCGGGAAAGCGCCGCCCTCCCGCC 60  
1 TGGCGGGGGGAGAGAGACCTTACCGCGCTCTGCTCCGAGGCTCTTGTGAGAGAGACC 120  
61 TGGCGGGGGGAGAGAGACCTTACCGCGCTCTGCTCCGAGGCTCTTGTGAGAGAGACC 120  
61 TGGCGGGGGGAGAGAGACCTTACCGCGCTCTGCTCCGAGGCTCTTGTGAGAGAGACC 120  
121 CGGGTGGAGACAGCCCTCCCTATTACCGCCGCTTTCTGGAGCGCTTTCCACCTGAAG 180  
121 CGGGTGGAGACAGCCCTCCCTATTACCGCCGCTTTCTGGAGCGCTTTCCACCTGAAG 180  
121 CGGGTGGAGACAGCCCTCCCTATTACCGCCGCTTTCTGGAGCGCTTTCCACCTGAAG 180  
181 GCCCTGGCGCGGCGCTTCCCTGAGAGAGCTTTAGGGGTGGCGAGGGGGGGGGCTACTAC 240  
181 GCCCTGGCGCGGCGCTTCCCTGAGAGAGCTTTAGGGGTGGCGAGGGGGGGGGCTACTAC 240  
181 GCCCTGGCGCGGCGCTTCCCTGAGAGAGCTTTAGGGGTGGCGAGGGGGGGGGCTACTAC 240  
241 CGGCGGGGCGGAACACCTCCACCGCGCTGCGCGGAGAGAGAGAGCTTCCCGAGCTTC 300  
241 CGGCGGGGCGGAACACCTCCACCGCGCTGCGCGGAGAGAGAGAGCTTCCCGAGCTTC 300  
241 CGGCGGGGCGGAACACCTCCACCGCGCTGCGCGGAGAGAGAGAGCTTCCCGAGCTTC 300  
301 GCCGAGCTTGGGGGCTTCTGCTGCTGGGCGCTTACACCGCGCGGCTGCGCTTCATC 360  
301 GCCGAGCTTGGGGGCTTCTGCTGCTGGGCGCTTACACCGCGCGGCTGCGCTTCATC 360  
361 GCCCTGGCGGAGAGGGGTGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
361 GCCCTGGCGGAGAGGGGTGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
421 TTGCGCCGGGAAAGCCCAAG 480  
421 TTGCGCCGGGAAAGCCCAAG 480  
421 TTGCGCCGGGAAAGCCCAAG 480  
481 GAGGGGCTGGAGACCGGGGGGTGGAGAACAGGCGCTTATGAGGTCCGGGGCGACGCTTC 540  
481 GAGGGGCTGGAGACCGGGGGGTGGAGAACAGGCGCTTATGAGGTCCGGGGCGACGCTTC 540  
481 GAGGGGCTGGAGACCGGGGGGTGGAGAACAGGCGCTTATGAGGTCCGGGGCGACGCTTC 540  
541 CTGCGGAAAGCGGCGGCTGGGGGGGTGGAGAACAGGCGCTTATGAGGTCCGGGGCGAG 600  
541 CTGCGGAAAGCGGCGGCTGGGGGGGTGGAGAACAGGCGCTTATGAGGTCCGGGGCGAG 600  
541 CTGCGGAAAGCGGCGGCTGGGGGGGTGGAGAACAGGCGCTTATGAGGTCCGGGGCGAG 600  
601 GCCCGGCGGCGCTACCCCGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
601 GCCCGGCGGCGCTACCCCGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
601 GCCCGGCGGCGCTACCCCGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
661 CTGCTCTCTCTGGGGGAG 720  
661 CTGCTCTCTCTGGGGGAG 720  
661 CTGCTCTCTCTGGGGGAG 720  
721 CTCTACGCGGCGCTCTTCCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
721 CTCTACGCGGCGCTCTTCCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780

Db 721 CTCTACGCGGCGCTCTTCCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
Qy 781 GTGAGGCTTAGGCGGAG 840  
Db 781 GTGAGGCTTAGGCGGAG 840  
Qy 841 GAGGTGGGGGGGCGCTTTGGGAAAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
Db 841 GAGGTGGGGGGGCGCTTTGGGAAAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
Qy 901 AAGCTCATGAG 960  
Db 901 AAGCTCATGAG 960  
Qy 961 CCCCTCCCGAGCGCA 975  
Db 961 CCCCTCCCGAGCGCA 975

RESULT 2  
AF377342 1089 bp DNA linear BCT 02-NOV-2001  
LOCUS Deinococcus radiodurans A/G-specific adenine glycosylase gene,  
DEFINITION partial cds.  
ACCESSION AF377342  
VERSION AF377342.1 GI:16588987  
KEYWORDS  
SOURCE  
ORGANISM  
Deinococcus radiodurans R1  
Deinococcus radiodurans R1  
Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
Deinococcaceae; Deinococcus.

REFERENCE  
AUTHORS  
TITLE  
Li, X. and Lu, A.-L.  
Molecular cloning and functional analysis of the MutY homolog of  
Deinococcus radiodurans  
J. Bacteriol. 183 (21), 6151-6158 (2001)  
MEDLINE  
21475757  
PUBMED  
11591657  
REFERENCE  
AUTHORS  
TITLE  
Li, X. and Lu, A.-L.  
Direct Submission  
Submitted (04-MAY-2001) Biochemistry and Molecular Biology,  
University of Maryland, Baltimore, 108 North Greene Street,  
Baltimore, MD 21201, USA

FEATURES  
source  
1..1089  
Location/Qualifiers

CDS  
1..1089  
/organism="Deinococcus radiodurans R1"  
/mol\_type="genomic DNA"  
/strain="R1"  
/db\_xref="taxon:243230"  
1..>1089  
/note="similar to MutY"  
/product="A/G-specific adenine glycosylase"  
/protein\_id="AA126976.1"  
/translation="MTLPVSASGPAPSPPEVGAIRRLDLSGFDAAGRLPRLDEGR  
RDPRVVAEILLOOTVARGLYERLEAFPTVQALAAAPQAVLKAWSGGTYAR  
ARNLHRAAIIIDEGFPQDYAGMLAPGVGYTAAVSSLAGBRVANDNVRVLS  
RLAEHRSKDVNOADRLDPAAPGAMVANDLGAITICVPSKSPACDRCPVSHCA  
AVOICQPDPPAPRAPQAREVRAVALLIGAEYAVLEKREGSLIGLFGPLIEIGA  
RETAADLARIQALRGAEVKEGCTGTVGHTHRRLSVETVAEADRPQVRGALSR  
LDHKLVDWRQRLFGAALDSSPLGA"

ORIGIN

Query Match 18.1%; Score 176.4; DB 1; Length 1089;  
Best Local Similarity 58.4%; Pred. No. 1.9e-14;  
Matches 391; Conservative 0; Mismatches 261; Indels 18; Gaps 4;  
Qy 66 GGGGAG 125  
Db 126 GGGGCGGGCGAGCCTTACCGGGGTGGGTGGCGAGATTGCTGCAACAGACGAGAGT 185

QY 126 GAGAGAGCCCTCCCTATTACCGCCGCTTTCTGAGAGCCCTTCCACCTGAAGGCCCT 185  
 Db 186 GGGCGGGGGGCTGGGACTACGAGCGCTTCTGAGAGCTTTTCCGAGGGGCGGCT 245  
 QY 186 GGGCGGGGCTTCCCTGAGAGAGCTTGAAGGCTTGAGAGGGGGGCGGCTACTACCGGCG 245  
 Db 246 GGGCGGGGCGGCGGAGAGAGCGGCTGCTCAAGGCTTGAGAGGTTGTGCTACTACCGGCG 305  
 QY 246 GGGCGAGACACTCCACCGCCCTGCGCCGAGAGGCTGAGAGAGC-----TTCCCGGAGCTT 299  
 Db 306 TGGCCCGACCTGACCGGGGAGCGGCGCATCATGACGAGAGAGGGGTTTCCGCAAGCTA 365  
 QY 300 GCGCGAGCTTGGGGGGCTTCTGCTCTGGGCTTACACCGCGCGCGGCTGCTCAT 359  
 Db 366 GCGCGGGTGGCTGGCTTGGCCCGGCTGAGGCGCTTACACCGCGCGGCTGAGAGCACT 425  
 QY 360 GCGCTTGGGGAGAGGGGTGGGCGGCTGAGAGGAGCGTCCGAGGCTCTCCGCGCT 419  
 Db 426 GCGGCTGGGAGAGCGGAGCGGCTGAGAGCGAGCGGAGCGGCTGCTGCGCGCT 485  
 QY 420 CTTCGCGCGGAGAAAGCCCAAGAGAGAGCTTTTCCGCTCGCCCGAGGCTCTCC 479  
 Db 486 GCGGCGAGAGGCTACCGCTCCGCAATGGGTGACAGAGAGC---TGACGCTTGGCT 542  
 QY 480 GAGAGGCTGAGACCGGGGGGTGTGAAACAGGCTTATGAGCTCGGGGCGACGCTGTG 539  
 Db 543 GAGACCGGCTGAGACCGGCTGCTGAAAGAGCGCTATGAGCTCGGGGCGACCATTTG 602  
 QY 540 CCGGCGAAAGCGGCGGCTTGGGCGGCTGCGCCCTGAGGGGCTTCTGCGGGGAGAA--- 596  
 Db 603 GGTGCGCAAGTCCGCGGCGGTGACCGCTGCGGCTGAGCGCCACCTGTGCGGCTACCA 662  
 QY 597 ---GAGAGCCCGCGGCGCTACCCCGCGCCAGAGAGCG---CCGGCGAGAGAGAGAGCG 650  
 Db 663 ACTGGGTCAGACCGGCTGACTTTCTGCGCCCAAGCGCGCGCGCGAGGTCG 722  
 QY 651 CTTGCTGCTGCTCTCTCTCGGCGGAGAGGGGTGACCTGAGAAAGCTTGAAGGCG 710  
 Db 723 GCGGCTGCGCTCTCTGATCGGAGAGCGCGAGTACCGCTGTGAGAAAGCGGAGGCTC 782  
 QY 711 CTTCCAGGCG 720  
 Db 783 GCTGCTGGGC 792

RESULT 3  
 AB002060/c 11378 bp DNA linear BCT 22-NOV-1999  
 LOCUS Deinococcus radiodurans R1 section 197 of 229 of the complete  
 DEFINITION chromosome 1.  
 ACCESSION AB002060 AB000513  
 VERSION AB002060.1 GI:6460082  
 SOURCE Deinococcus radiodurans R1  
 ORGANISM Deinococcus radiodurans R1  
 Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
 Deinococcaceae; Deinococcus.  
 REFERENCE 1 (bases 1 to 11378)  
 AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,  
 Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,  
 Mofat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,  
 Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,  
 Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.  
 TITLE Genome sequence of the radioresistant bacterium Deinococcus  
 radiodurans R1  
 JOURNAL Science 286 (5444), 1571-1577 (1999)  
 MEDLINE 20036896  
 PUBMED 10567266  
 REFERENCES 2 (bases 1 to 11378)  
 AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,  
 Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,  
 Mofat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,  
 Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,  
 Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.

TITLE  
 JOURNAL  
 Direct Submission  
 Submitted (08-NOV-1999) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA  
 FEATURES  
 source  
 1. 11378  
 /organism="Deinococcus radiodurans R1"  
 /mol\_type="genomic DNA"  
 /strain="R1"  
 /db\_xref="taxon:243230"  
 /chromosome="1"  
 118. 879  
 /gene="DR2274"  
 118. 879  
 /note="similar to PID:1001216 PID:1001276 percent  
 identity: 51.02; identified by sequence similarity;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAPI1822.1"  
 /db\_xref="GI:6460083"  
 /translation="MTEQAQYRDPRLVLYIDRLNRGADDFPLANETPGCRLLDL  
 GCGTGRITLALAGQVGTGLPAPASLEVAQRPKAEACWLGSGQADAPTAFLDC  
 LMTAHVAQVVEDDAMQDVLQIHRLVPGGRSLSPDMRDPAAAMDYDSDAGKERTLT  
 LLDGSEAEETCDVLDVAGPVLVSGPVHFAETRELPQDVLVTSRLRFRREDLRAS  
 LQAGFAVEQVYGVGHEPVGTCGELVVAARRPG"  
 840. 3032  
 /gene="DR2275"  
 840. 3032  
 /note="similar to SP:056243 PID:1325924 percent identity:  
 79.97; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="exonuclease ABC, subunit B"  
 /protein\_id="AAPI1823.1"  
 /db\_xref="GI:6460084"  
 /translation="MRTTGRGAGQAGVTRLPRLPSSLRGCRKLLADVOTLEPRLP  
 CVCLARCRRTLEVVMKARSEKPSGSDOTPLASLYDGLSGRLVTLGATGTGPT  
 YSMAKVIEVQRPALINAPNKILTAQLASFEPFPAVAEFLISTDYIOPETAVFG  
 KDLFIERDASVNOEIERLRHSTRLSLTRDITIVASVSCIYGGDKETALNAIVK  
 KGVMPREDDELGRVLMQYERNDIELMPGFSGVGEVSTYGADEPKETALNAIVK  
 ERISVHPLTGERLGDLDATVVYPAKHVVSAGNIERAGSIOELDERLEYFHSVGK  
 LLEAQLKERITLDLEMLKVLGYCSGIEVNSRHIDGAPGATPYTMIDYRDPDITFI  
 DESHATVPQIGMANGDRAKQTLVDYGFRLPEAMDRPLNFEFSLSKTQYFVSAT  
 PGPFENQVSDSVADQIIRPTGLADPEVTRIPGQILDLGRVREKRTAAGEKRLVTL  
 TKRMSDELTEYLEKGVKARYLHSDISVERQVIIRDLRGLGHVGLVIGLIRGLDL  
 PEVSLVALIDADKGLFASDRALIGTIGRAARVNGEVLVYADVTPAMFAMDETRAR  
 RREKSLAVNEHEGITPRTVIGKVRDVRGELEGEISSEVSGRDLTQLTDLIED  
 MNGASEMDPEKASLRDQIRALIAKLGQEFQCATYPOKARRKGR"  
 3035. 3184  
 /gene="DR2276"  
 3035. 3184  
 /note="identified by Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAPI1832.1"  
 /db\_xref="GI:6460093"  
 /translation="MGRGAKMLPAALJALIFLVVALVWMTGPTAVVLAFLPLAAVLS  
 ASGRC"  
 complement (3287. 4006)  
 /gene="DR2277"  
 complement (3287. 4006)  
 /gene="DR2277"  
 /note="similar to GB:AB000782 percent identity: 66.67;  
 identified by sequence similarity; putative"

gene  
 CDS

gene  
CDS  
/codon\_start=1  
/transl\_table=11  
/product="amino acid ABC transporter, permease protein"  
/protein\_id="AAFI1824.1"  
/db\_xref="GI:6460085"  
/translation="MSRSKPGGMDLTGFTQALSGDYPQLLSGLGTLAVSICALV  
VSIVVCTGLIVRYVRPLDPLGIVYVGGIPLVILSVIFGFLPTLGLIAGFP  
AAVLAIGLSAATSEIVRGSLGAGIAGSLRGALRFLVPLQAMVLAIP  
ALGNEFVSLILSSLASAVTLDLFSQGRITNATYRQFEYVAIVLAVLFTLTLR  
AVRALEHRLSRQSLPERRVI"  
/complement(3993..4811)  
/gene="DR2278"  
/note="similar to SP:P10344 PID:41569 GB:U00096  
PID:1651365 PID:1651371 percent identity: 53.85;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="amino acid ABC transporter, periplasmic amino  
acid-binding protein"  
/protein\_id="AAFI1825.1"  
/db\_xref="GI:6460086"  
/translation="MGLVRNVRGRILNRYSGMKRLSAILLVLAISCSAGASYAY  
KKKGLVLGDPFAPFEFKSGDGVQGDIDIAVAHADLVLRIRVPGALMPO  
AATSGRVVAMSGITITPERATVTFPSQYSAQVIVRSNGPKFTMPADVKKT  
GVQANTGOVAACQLKPRGATIKYVDPAAGLADVRSRIALIGDAFTVDLKKRL  
PGQPDKAGNVAVEDYGMVFRKGSDLAAANRTLAALKRSGEYOKLNMWVQK"  
4846..6141  
/gene="DR2279"  
4846..6141  
/gene="DR2279"  
/note="similar to GB:AL009126 percent identity: 59.28;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="alcohol dehydrogenase, zinc-containing"  
/protein\_id="AAFI1826.1"  
/db\_xref="GI:6460087"  
/translation="MGRKCKGWPVKVFLHSLSPGAGSTKAVVHMGIDRLDVP  
EPKTEAPDAIVRLTASALIGTDLHFRGMSDMVPTLIGHGAVIIVAVGPEVNF  
VPGDRIYPSASGCGYCPREBGTNSQDNANPNFGAGTAFVGGVAVVGLGCGV  
KVRVYVNSSLVKLIDVSDDOAIMSDIFPTAYFGAELGVGAVVVLVGGPAG  
FSIISALGLATRIAVDRUDRLDMKRGKAVINREKEDPUEAKITGLGLADVC  
IDAVGVQAHNAHAPAKPDAKEKQFEQVEVAIPDKERDQMTPGDAPSAASMA  
IEMTKAGQIGIIGVSPQMTYPIGKANNKILTRNGNCNHRALIIPLVDVAAGV  
DPAVLTEHHLGSAIEFAFEDKRPQGMIKVLELPQAG"  
complement(6214..7704)  
/gene="DR2280"  
complement(6214..7704)  
/gene="DR2280"  
/note="similar to GB:X76083 SP:P41006 PID:43121 percent  
identity: 70.23; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="uracil permease"  
/protein\_id="AAFI1827.1"  
/db\_xref="GI:6460088"  
/translation="MHLAVMARAGQSSSSAWGLDARDFYREARALRARRPMSATL  
RWTMSSEKATAEACFCSPSPAHYTPMTQAPLPIPLPQRRVVLGLQHSIM  
RGSTIVLVIVLQSPVALFAGVATLIFHLTRGQVIFGSSFAFLAPALVIKEM  
GFAAAGGILAAGLMYVPSGLYKIFGRERILRRPPVPTGVIIVGLGASVAVQ  
AKTNGLALVTLAIIASITGRGLFRMTPLILGVYGVYVALATGGVAGLADIAA  
APLGLPPEHGPALDMRAVAIIAPVAIVTFEHDVAVNRRVYKFNALPGLSRTL  
LADGIANSSAVMGPAATVAENTGVALTRVDPVIOGAVFALLFGSPKLAAY  
LOGLPQGVGVSLIFGMIASVIGIRTLAEQVDFAHSRNLIIVSLIIVLGLGASFP  
LNVAGTKLEIHMALLALVGIIVANLLRPRPEISKADGPGVG"  
7746..8240  
/gene="DR2281"  
7746..8240  
/gene="DR2281"  
/note="identified by Glimmer2; putative"

gene  
CDS  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAFI1833.1"  
/db\_xref="GI:6460094"  
/translation="MGAGQKAKAEVLTPPPPOSRTLYPMNAKGDLLRLVSLGCTP  
TEEAHGPHERTHVYKWSGVIAAEGRTKDARLAEALRLDSDAPVP  
AAPATPIAQSSPWPVIYQVLAEAAMEFARBDATLDEVARDAGRFREBLADLGH  
GPA"  
8243..8614  
/gene="DR2282"  
8243..8614  
/gene="DR2282"  
/note="similar to SP:P45465 PID:606088 GB:U00096  
PID:1789538 percent identity: 56.57; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAFI1828.1"  
/db\_xref="GI:6460089"  
/translation="MPTSRPRGHOGAABARAAALHESLGREIVARNYRIPGEIDL  
VSRPGTLVTFEVRORARYSALDSVTPKRLMHRAALEYLTRCGRDDLPCLRL  
EVLIRGEADVTGELRLAVEG"  
complement(8625..9431)  
/gene="DR2283"  
complement(8625..9431)  
/gene="DR2283"  
/note="similar to SP:Q55282 percent identity: 66.27;  
identified by sequence similarity; putative"  
Query Match 18.1%; Score 176.4; DB 1; Length 11378;  
Best Local Similarity 58.4%; Pred. No. 9.1e-15;  
Matches 391; Conservative 0; Mismatches 261; Indels 18; Gaps 4;  
gene  
CDS  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAFI1833.1"  
/db\_xref="GI:6460094"  
/translation="MGAGQKAKAEVLTPPPPOSRTLYPMNAKGDLLRLVSLGCTP  
TEEAHGPHERTHVYKWSGVIAAEGRTKDARLAEALRLDSDAPVP  
AAPATPIAQSSPWPVIYQVLAEAAMEFARBDATLDEVARDAGRFREBLADLGH  
GPA"  
8243..8614  
/gene="DR2282"  
8243..8614  
/gene="DR2282"  
/note="similar to SP:P45465 PID:606088 GB:U00096  
PID:1789538 percent identity: 56.57; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAFI1828.1"  
/db\_xref="GI:6460089"  
/translation="MPTSRPRGHOGAABARAAALHESLGREIVARNYRIPGEIDL  
VSRPGTLVTFEVRORARYSALDSVTPKRLMHRAALEYLTRCGRDDLPCLRL  
EVLIRGEADVTGELRLAVEG"  
complement(8625..9431)  
/gene="DR2283"  
complement(8625..9431)  
/gene="DR2283"  
/note="similar to SP:Q55282 percent identity: 66.27;  
identified by sequence similarity; putative"

Oy 651 CCTGTCGCTCTCTCTCTCGGGCGGAGGGGTGCACTTGAGAAAGCTTGAGGGCG 710  
Db 10568 GGGCGTGGCTCTCTCTGATCGGGACGCCAGTAAGCCCGTCTGAGAAAGCGAGGGCTC 10509  
Oy 711 CTTCAGGGC 720  
Db 10508 GCTGCTGGCG 10499

RESULT 4  
AF543480 8249 bp DNA linear BCT 07-MAR-2003  
LOCUS  
DEFINITION Uncultured bacterium plasmid pAK106 Soda (soda) gene, partial cds;  
Grea (grea), Muty (muty), putative aminopeptidase, and putative  
response regulator genes, complete cds, and unknown genes.  
AF543480  
ACCESSION AF543480.1 GI:28875482  
VERSION  
KEYWORDS  
SOURCE uncultured bacterium  
ORGANISM Bacteria; environmental samples.  
REFERENCE 1 (bases 1 to 8249)  
AUTHORS Knietsch, A., Waschewitz, T., Bowien, S., Henne, A. and Daniel, R.  
TITLE Construction and Screening of Metagenomic Libraries Derived from  
Enrichment Cultures: Generation of a Gene Bank for Genes Conferring  
Alcohol Oxidoreductase Activity on *Escherichia coli*  
Appl. Environ. Microbiol. 69 (3), 1408-1416 (2003)  
12620823  
2 (bases 1 to 8249)  
REFERENCE Knietsch, A. and Daniel, R.  
AUTHORS Direct Submission  
TITLE Submitted (06-SEP-2002) Institut f. Mikrobiologie, University of  
JOURNAL Goettingen, Grisebachstr. 8, Goettingen D-37077, Germany  
JOURNAL Location/Qualifiers  
FEATURES  
source 1. .8249  
/organism="uncultured bacterium"  
/mol\_type="genomic DNA"  
/isolation\_source="soil"  
/db\_xref="taxon:77133"  
/environmental\_sample  
/plasmid="pAK106"  
/note="anonymous microbial sequence conferring formation  
of carbonyls from short-chain polyols on *Escherichia coli*"  
1. .618  
/gene="soda"  
/note="putative superoxide dismutase"  
/codon\_start=1  
/transl\_table=1  
/product="Soda"  
/protein\_id="AA059964.1"  
/db\_xref="GI:28875483"  
/translation="DLAYRBEFRFPPEAGLSATLATHIELYRGVVAQRELLGL  
ADBLRDEAGVTRPREAVSRKMSFEINGLVHWEYEQLAGRSGRLVPPSSVAAGS  
MDVAFGGFDGRADVALELATTRGIVVAVADASQGLVINMADLHLLLPAGHIVF  
VLDMERAYMDPFGKRGAVSKFILLNTWEEVVEARFARBNGA"  
complement (673. .2466)  
/gene="grea"  
/note="grea"  
/complement (673. .2466)  
/gene="grea"  
/note="putative transcription factor"  
/codon\_start=1  
/transl\_table=1  
/product="Grea"  
/protein\_id="AA059965.1"  
/db\_xref="GI:28875484"  
/translation="MRHADGLOGASALRPAOGADAGAOVHQAIVGRDALRGQGGIG  
LRPORTVCGREVAVAKTECGHPFDVALEDGPIAEERGDGRCGAPTCGRGQH  
LAARBELAAMHHNRRTAVOVAARAVALPREHVVORRGORLHVGTTLGEPV  
VVDHGHLLIOLHILGQPHAVRVICALPGOVAMLLPLPAHNAAGALOLAHSSGVSR  
RGARWRPGPCTRDVFGGGHQAIVHLLQGFAPLARFLGDALLOYTRGDLALQCLAR

gene  
CDS  
1834. .2877  
/gene="muty"  
1834. .2877  
/note="muty"  
/note="putative adenine glycoylase"  
/codon\_start=1  
/transl\_table=1  
/product="muty"  
/protein\_id="AA059966.1"  
/db\_xref="GI:28875485"  
/translation="MKLESLASRYRWQROHGRHGLPMQGTODPYRVMLEVMLOQT  
VVIVIDYARFLORFPDVQALAAAPLDVDLALMSGLVGRARHLRCAQAVVQHG  
EPPSSSEMLTLPEIGRSTAAATAAFCGRRVAILDGNVRLTRALGFGDLSRPN  
ERALMAQOTLLPAQGITAYTQGLMDGAGICTLRPOCAACPLQPCVVAHOGRPA  
YVKTRLKRGQRAVVMVLMHGLQVLRPREQGVAGLMSLPENADSLGTLAA  
WPGHEWLPVPALTHFDWTLHPLRWQLPARVPAARRKAVEQALPTGRWFTQSLA  
ASLPAPRYKLKAG"  
complement (2904. .3335)  
/note="orf6; hypothetical protein"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=1  
/product="unknown"  
/protein\_id="AA059967.1"  
/db\_xref="GI:28875486"  
/translation="MKRTTFLTAIALSAGSACVPAQTDTGGLLDKAKKTYLSD  
KDVAESRCFNDCAVWPPFMAVDGAAGKLTIVARKGGQQAMWGMPLYTAGDA  
KVDTSGDGSGGWVHVIKAGGDKPEKRAATGNAADPYK"  
3742. .4827  
/note="orf7; hypothetical protein"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=1  
/product="unknown"  
/protein\_id="AA059968.1"  
/db\_xref="GI:28875487"  
/translation="MRHGCSAAHCGASSSSSAVMPAHRVAVACOARLGSAGAGSSP  
TSSRADAAATPPVYSKQAGPVRFGDQAALESFATQVAVVAASGMAAAGSGRRAD  
HSPVASASHPRSVCVAAAWATQARRPRTSREGAPSSARPVQDCAPIDVA  
FSHAGPALRLRAPRGVWRIRIDQAPAGQPGFSGFKHPAAGCNVIVIGPAVERQ  
DRLPVLRQPPPTAQRQAAGSGSLSPMSQVWRDAGIVSAQPSRSGSIVRSARR  
ARRRSMARLTCCOPGHRPCDREKQABERGFLLVGPARGAPAGEWVACRSRMSRRRA  
WPAWSRSGRPVRRPCSRRLRRPWR"  
complement (5521. .5868)  
/note="orf8"  
/codon\_start=1  
/transl\_table=1  
/product="putative aminopeptidase"  
/protein\_id="AA059969.1"  
/db\_xref="GI:28875488"  
/translation="MNALVETDSDGLNDIRGQVTPAQVVDALRAARSGFVAESVG  
AGCTGTAFFKMGKIGITSSRAVTAGSSSYTVGLVQNSYGDVLWAGVPVGAASSPPGGA  
RTRVPRPAPARA"  
5874. .6179  
/note="orf9"  
/codon\_start=1  
/transl\_table=1  
/product="putative response regulator"  
/protein\_id="AA059970.1"  
/db\_xref="GI:28875489"  
/translation="MGCPILLDTVARRAQNGERPIDLAERNAVLTILARVEKIVSK  
EATIQAVAMGSEELTPALIEVYSRLRAKLEPAGVKIRTVRGFIMLEFREKKAQR"  
complement (6591. .7028)  
/note="orf10; hypothetical protein"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=1

```

/product="unknown"
/protein_id="AAO59971.1"
/db_xref="GI:28875490"
/translation="MRRPRTSSAREGCGKSGHTPAVVEALGALPDVEVTTAMGQPLAKVGNKFEVCVAHSKSLKSPNSLYMMAFADRLALVDDPRTVYLRGHYLNHPCVLRLIRLRDLALRDLVVGAFHLVAVQKSSSDIGSQGSRPLRDR"
complement(7784..>8249)

```

```

/product="unknown"
/protein_id="AA059972.1"
/db_xref="GI:28815491"
/translation="IGRRARPDAGSGIAA'PPEDAWVSRRSGRRDTRPTGSCVATLGDG
RQHRRAAVATPPGSPESHQPOA'PDAGASGPRRAE'ENHAGGADDEDTDRD'LAAPV
NARGGAASGGRRRRGRDRDHLAAGDA'EHHPRSHSLA'LOGRGRRGQVNSGE"

```

ORIGIN

Query Match	16.7%	Score 163.2	DB 1	Length 8249
Best Local Similarity	54.7%	Pred No. 5.8e-13		
Matches 408		0	Mismatches 308	Indels 30
				Gaps 3

OY 2 TGAAGAGCCTGGACGGAAAGCCCTCCGCTGGCTAGTACCGGGAAAGACCCCGGCC---CTCC 58  
 Db 1841 TGAAGAGCTGGACCGCCCGCTGGCTGGCTGGACCGGCAACATGACCGCCACCGGGCTGC 1900  
 OY 59 CCTGACGGGGGAGAGAGAGACCTTACCGCGATCTGATCTCCGAGGCTCTTCTGACAGAGA 118  
 Db 1901 CCTGGACAGGGGACACAAAGACCCGTATCCGCGATGTGGCTGTCCGAGGTATGCTGCAGCAGA 1960  
 OY 119 CCGGGGTGAGAGAGGCCCTTCCCCATTATACCGCCGCTTCTGTGAGCGCTTTTCCACCTCGA 178  
 Db 1961 CGCAGGTGTGATCACCTGTATCATCTATACGCCCGGGTCTCTCAAGTTTCCGACGTGC 2020  
 OY 179 AGGCGCTGGCCGCGGCTTCCGTGAGAGAGTCTTAAAGGATCTGGCAGGAGGCGGGCTACT 238  
 Db 2021 AGGCGCTGACCGGCGCGCGCTGTGACAGACGTCTCGCGCTCTGAGACGCGCTGGGCTACT 2080  
 OY 239 ACCGCGGGGCGGAAACCTTCCACCGCGCTGGCCCGAAGCGTGGAG-----G 283  
 Db 2081 ACAGCGGCGCGCGCACCTGTGACCGCTGTGGCGCAGCGAGTGGTGCAGATGCGCGCG 2140  
 OY 284 AGCTTCCCCAGCTTTCGCCCAAGCTTCCGGGGCTTCTGATGCTCGGGGCTTAAACCGGCG 343  
 Db 2141 AGTTCCCCCGCAGCGACGAGATGTGGGCCACCTGCTCCGATATCGGCGCTTCACCGCGG 2200  
 OY 344 CGGGCGGTGAGCTTCATCGGCTTCCGGGAGCGGGTGGCGGCGGTGACCGGAACTGCCGA 403  
 Db 2201 CAGCCATGCGCGCTTCTGCTTCCGCGAGCGGGTGGCCATCTCGATGGCAACGTCAAC 2260  
 OY 404 GGGTCTCTCCGCGCTCTTTCGCGCCGGGAAAGCCCAAGAGAGAGCTTTTCGCTTCG 463  
 Db 2261 GGGTATTGACCGGAGACATCGGTTTGGCGAGCATCTTCGCGGCCCGCAAAAGAACGTG 2320  
 OY 464 CCCAGGCGCTCTCCCCGAGGCGGTGACCCCGGGGTGTGAAAC-----CAGG 511  
 Db 2321 CGCTGTGGGCGCAGGCGCCAGACCTTGTGTGCCTCGCGCAGGAGGATCAACGCTACACGCAAG 2380  
 OY 512 CCTCATGAGAGCTCGGGGCGCAGGTCTGTGCTCGCAGAAACGCGCCCGTTGCGGGGCGTCC 571  
 Db 2381 GCTTATGAGACTTGGGCGCGCGGCACTGTGACCTTCGCGCGCGCAGATGCGGTGCTGCC 2440  
 OY 572 CCTTAGGGGCTTTCGCGGGGAGAGAGAGGCCCGCGGCGTAAACCCCGCGCCAGAAAGC 631  
 Db 2441 CCTTGAAACCGGTCTGCGGTGGCGATGCGCAGAGAGCGGCGCAGAGGCTTACCGGTGAAGA 2500  
 OY 632 GCGGGCGAAGAGAGCGCTGTGCGCTGTCTCTCTCGGCGGGAAGGGGGGTGACCC 691  
 Db 2501 CACGCGCTCTCAAAACGTGGGCAAGCGGCGCAAGGTGTGTTGTGGCTGTGCACGGCTGCG 2560  
 OY 692 TGAAGAGGCTTGAAGGGGCGCTTCCAG 717  
 Db 2561 AGGTCTGCTCTGATGCTCGGCGGAG 2586

RESULT 5  
AE016923/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AUTHORS  
CONSRPM  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

303642 bp DNA linear BCT 14-OCT-2003  
AE016923 Chromobacterium violaceum ATCC 12472 section 14 of 16 of the complete genome.  
AE016923 AE016825  
AE016923.1 GI:34104942  
Chromobacterium violaceum ATCC 12472  
Chromobacterium violaceum ATCC 12472  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Chromobacterium.  
1 (bases 1 to 303642)  
Brazilian National Genome Project Consortium  
The complete genome sequence of Chromobacterium violaceum reveals remarkable and exploitable bacterial adaptability  
Proc. Natl. Acad. Sci. U.S.A. 100 (20), 11660-11665 (2003)  
22882880  
14500782  
2 (bases 1 to 303642)  
Vasconcelos A.T.R., de Almeida, D.F., Almeida, F.C., de Almeida, F.M.

CONSPIRM TITLE	JOURNAL	MEDLINE PUBMED	REFERENCE
Brazilian National Genome Project Consortium The complete genome sequence of <i>Chromobacterium violaceum</i> reveals remarkable and exploitable bacterial adaptability	Proc. Natl. Acad. Sci. U.S.A.	100 (20), 11660-11665 (2003)	
2 (bases 1 to 303642)			
Almeida, L.G.P., de Almeida, R., Gonçalves, J.A.A., Andrade, E.M., Antônio, R.V., Araújo, J., de Araújo, M.F.F., Filho, S.A., Azevedo, V., Batista, A.J., Bataus, L.A.M., Batista, J.S., Belo, A., Vander Berg, C., Blamey, J., Bogo, M., Bonato, S., Bordignon, J., Brito, C.A., Broccchi, M., Burity, H.A., Camargo, A.A., Cardoso, D.D.P., Carnieli, N.P., Carraro, D.M., Carvalho, C.M.B., Cascardo, J.C.M., Cavada, B.S., Chneiweil, M.O., Páta, T.B.C., Duran, N., Fagundes, N., Falcão, C.L., Fantinatti, F., Farias, I.P., Felipe, M.S.S., Ferrari, L.P., Ferto, J.A., Ferro, M.I.T., Franco, G.R., Freitas, N.S.A., Furlan, L.R., Gazzinelli, R.T., Gomes, E.A., Gonçalves, P.R., Grangeiro, T.B., Gratiapaglia, D., Grisard, E.C., Guimarães, C.T., Hanna, E.S., Hungria, M., Jardim, S.N., Laurino, J., Leal, L.C.T., Passarella, L., Lima, A., Loureiro, M.F., Lyra, M.C.P., Macedo, M., Madeira, H.M.F., Manfio, G.P., Maranhão, A.O., Martins, M.S., di Mauro, S.M.Z., de Medeiros, S.R.B., Meissner, R.V., Menck, C.F.M., Moreira, M.A.M., Nascimento, F.F., Nicolas, M.F., Oliveira, J.G., Oliveira, S.C., Palhao, R.F.C., Parente, J.A., Pedrosa, F.O., Pena, S.J.D., Pereira, J.O., Pereira, M., Pinto, L.S.R.C., Pinto, L.S., Porto, J.I.R., Portich, D.P., Nero, C.E.R., Reis, A.M.M., Rigão, U., Rondinelli, E., dos Santos, E.B.P., Santos, F.R., Schneider, M.P.C., Seunarez, H.N., Silva, A.M.R., da Silva, A.L.C., Silva, D.W., Silva, R., Simoes, I.C., Simon, D., Soares, C.M.A., Soares, R.B.A., Souza, E.R., Souza, K.R.L., Souza, R.C., Steffens, M.B.R., Steindell, M., Teixeira, S.R., Uremwy, T.T., Vettore, A., Wasseem, R., Zaha, A. and Simpson, A.J.G.			
Direct Submission			
Submitted (22-JAN-2003) Labinfo, UNCC - Laboratório Nacional de Computação Científica, Rua Getúlio Vargas 333, Petrópolis, RJ 25651070, Brazil			
FEATURES			
source	Location/Qualifiers		
	1..303642		
	/organism="Chromobacterium violaceum ATCC 12472"		
	/mol_type="Genomic DNA"		
	/strain="ATCC 12472"		
	/db_xref="taxon:243365"		
	complement(78..527)		
gene	/gene="rp11"		
	/locus_tag="CV3637"		
	complement(78..527)		
CDS	/gene="rp11"		
	/locus_tag="CV3637"		
	/note="Identified by sequence similarity; putative; ORF located using Glimmer/Genemark/Blatx/COG0359"		
	/codon_start=1		
	/transl_table=1		
	/product="50S ribosomal protein L9"		
	/protein_id="AA061299.1"		
	/db_xref="GI:34104943"		
	/translation="MQIIILEKAVNIIQLGDVYVNVKNGFARNFIIPOGSKAKKRAEAL"		

KEPFDARRAELEKQKQELILADAKVRAEKNEAVITTAQAQAGCDGRLFGSVTNVDVAEAV  
TAEFGVQIKHEVRLNGLPEPKFAIGEYDIALHHDVVTPEIKIIVWGEA"  
gene complement(549. .776)  
/gene="rpsr"  
/locus\_tag="CV3638"  
complement(549. .776)  
CDS /gene="rpsr"  
/locus\_tag="CV3638"  
complement(549. .776)  
gene /locus\_tag="CV3638"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/COG0238"  
/codon\_start=1  
/transl\_table=11  
/product="30S ribosomal protein S18"  
/protein\_id="AAO61300.1"  
/db\_xref="GI:34104944"  
/translation="MAROLFKKKFCRFPAEGIKEDYKSYDLKDFIAENGKIIPAR  
ITGTARYOROLTTAIKARPLAFLPYTDQH"  
complement(782. .1081)  
CDS /gene="r18"  
/locus\_tag="CV3639"  
complement(782. .1081)  
gene /gene="r18"  
/locus\_tag="CV3639"  
complement(782. .1081)  
/locus\_tag="CV3639"  
/note="identified by sequence similarity; putative; ORF  
located using Genemark/Blastx/COG2965"  
/codon\_start=1  
/transl\_table=11  
/product="primosomal replication protein"  
/protein\_id="AAO61301.1"  
/db\_xref="GI:34104945"  
/translation="MQRNLVLTVTVDREDALRYTPAGLPPVEMMLKQSRQSGGFER  
DVACIOAVWGEAEAKKAGKAGLRGVMNVTGFLSQRSLRNPRLVLTNEYVEFVG"  
complement(1100. .1474)  
gene /gene="rpsf"  
/locus\_tag="CV3640"  
complement(1100. .1474)  
CDS /gene="rpsf"  
/locus\_tag="CV3640"  
/note="identified by sequence similarity; putative; ORF  
located using Glimmer/Genemark/Blastx/COG0360"  
/codon\_start=1  
/transl\_table=11  
/product="30S ribosomal protein S6"  
/protein\_id="AAO61302.1"  
/db\_xref="GI:34104946"  
/translation="MRHVEIVFIVPDSEOVPMIRYKGMVLAEGKIHRLEDKGR  
ROLAIPIQGLKHAHYVLMANVECOAETLAEIEAFKNDAYLRHLTIKDRASIDASP  
MDEKAKNILDQAPAEVEASA"  
complement(1631. .1927)  
CDS /locus\_tag="CV3641"  
complement(1631. .1927)  
gene /locus\_tag="CV3641"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/COG1937"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAO61303.1"  
/db\_xref="GI:34104947"  
/translation="MKKLAMTLLSSGAHAAPIPFETYYIKLHNGMLEAEIVSAGAP  
DVGSSNQRLASGVAVSTRQSLWLMANGQPTVTTLRDGVVVDIKRKKL"  
complement(1984. .2307)  
gene /locus\_tag="CV3642"  
complement(1984. .2307)  
CDS /locus\_tag="CV3642"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/COG1937"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAO61304.1"  
/db\_xref="GI:34104948"  
/translation="MDDTHSSQACCHBEAQERSVVOGKALLKRLARIEGQVIGIS  
GMIESRQVCVDVLTQAAVKSALDAVVALQLENNHKGCVARALQOGAGTAVTELIDV  
VKKVR"

gene complement (2356 .3852)  
locus\_tag="CV3643"  
complement (2356 .3852)  
CDS locus\_tag="CV3643"  
/note="Identified by sequence similarity; putative; ORF  
located using Glimmer/Blastx/COG2133/TC:9.A.10.1.1"

gene /codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA061305.1"  
/db\_xref="GI:34104949"  
/translation="MEMBRANEMEMBRANEALSPULRPAASASHAGCGMAGCA  
AELPAAALGLARGLQPLDKRLPLERKGAALRGCTTAAVSLPRLPGKATPEPAYN  
DSVPAILLEFESETLGLRFNDRLPPTVMHNGRLPIPSDQDNPHDPVPPGGGRYR  
FSLEDPSAGSYWHPHGHGTQAAYRGAGVVSRRNDPLAHLEPVLVSDLDL  
PDGKIANSADNRDEEGQFVLNMGAMOKIALGSGERQRMIMATSAKVILKLAP  
AHEWIVGTGDGSLIEKRPREVELLPFGARAEALVVRPAKGPAALVALPYARGKM  
GREGOQLPLDI EAQGMA TASPRLRIIDPFLIAGTVTRVSEDMANPAAMFLI  
NGKFTDNMRDFTGVQAVABEINQADNDHFPHLHGTPQVTTERHDGKRAESYL  
AMRVNVPPGQVRYLRFRQDMPGPMMFHCHLLEHDAQMGTLDVRA"  
complement (3982 .4773)  
/gene="subB"  
/locus\_tag="CV3644"  
complement (3982 .4773)  
CDS /gene="subB"  
/locus\_tag="CV3644"  
/note="Identified by sequence similarity; putative; ORF  
located using Glimmer/GemMark/Blastx/COG0483"  
/codon\_start=1  
/transl\_table=11  
/product="myo-inositol-1(or 4)-monophosphatase"  
/protein\_id="AA061306.1"  
/db\_xref="GI:34104950"  
/translation="MHPMVAVAKAARBAGSVTORASLNDAIRVEKKGNPTEVD  
RAAEAIINVILEAPYGHATLAEEGAKGCSSEYEWIIDPIDGTNNPILHGPQYS  
IANMKQIOOAVVDPENRDLFTASRGGAIFNSDRIVSKRFMNNECLITGFSPVS  
DOSLDDYLGMLXDVIAKTARVAREGASIDLNVACRGTEGE FELNLKPMDIAGSL  
IVDEAGGITVDPTGEEQGWESSGDIVAGNPVLAQLQLLAIPHVK"  
4946 .5680  
/locus\_tag="CV3645"  
4946 .5680  
CDS /locus\_tag="CV3645"  
/EC\_number="2.1.1.-"  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/COG0565"  
/codon\_start=1  
/transl\_table=11  
/product="probable RNA methyltransferase"  
/protein\_id="AA061307.1"  
/db\_xref="GI:34104951"  
/translation="MONIRVILVARNPNHGNIGSABAMTKMTGLSRLYVLPKFPEE  
ANLASGVDLENATNYGSLREALSDDVVYACLTSSRRELTPLSSTPOTPELLAR  
ARDDEYOALVFGENTFGLSIEVEOCRLVTTIGNDRTSYSLINAMAVOYTYBLEPHT  
GVDEYVIRAGEAATQSEVDWGCHLEQAWDELGYFORRNSERLMRMTMFRRAML  
REEEDILRGFYKQVORTASGRSGDAD"  
5736 .6581  
/locus\_tag="CV3646"  
5736 .6581  
CDS

Query Match 16.2%; Score 157.8; DB 1; Length 303642;  
Best Local Similarity 53.1%; Pred. No. 9.9e-13;  
Matches 451; Conservative 0; Mismatches 362; Indels 36; Gaps 4;

22 CTCCTCCGCTGTAACCGGAAAACGCCGCCCTCCCCCTGGGGGGGGAGAAGAACCTT 81  
Db CTGTGGCTTGGAACAACGGCGCATGCGCCGCAAGACTTGCCTTGGAGGTGAAGACCCG 66889

82 TACCGCGTCCGTGCTCCGAGGTCTCTTCTCACAGACCCGGGTGAGCAGCGCTCCCC 141  
Db TATCGGCTGTGCTGTACAGATCATCTGCACAGACCAGGTGAAGAGCGTCTGCAC 66882

142 TATTACGCGCTTCTTGAGCGCTTTCACACCTTAAGGCGCTGCGCGGCTTCCCTG 201



Db 66820 TATTACCCGCTTCTGCGCCGCTTCCCGACGTGGCCAGCTTGGCCGCCGCGCGGTG 66761  
Qy 202 GAAGAGTCTTAAAGGCTTGGCAGGGGCGGCTACTACCGCGCGGAGAACCTTCAC 261  
Db 66760 GACGACGTGTGGCGCACTGAGCGGCTGGCTACTACGACGGGGCGCACTCTGCAC 66701  
Qy 262 CGCCTGGCGCAAGCTGTGAGGA-----GCTTCCCCGAGCTTGGCGAG 306  
Db 66700 AAGCGGGGAAAGATGTATGATGCTTTGGCGGCGCTTTTCCCCCGAGCGGAGCGG 66641  
Qy 307 CTTCGGGGGCTTCTGCTGTGCTGGGCTTACACCGCGCGGCGGCTGAGCTCATGCGCTTC 366  
Db 66640 CTGAGAGCAGCTGCGCGCGGTGCGGCGCTCCACCGCGCGGATTTCCGCTTGGCTTC 66581  
Qy 367 GGGAGCGGGTGTGGCGGCTGAGCGGAACTCGGAGGGTCTTCCGCGCTTTCG-- 424  
Db 66580 GGGCGGCGGAGACCATCTGTGAGCGGCAATGTCAAAAGGGGTGTGGCGCGCTGCTTGGC 66521  
Qy 425 -----CCCGGAAAGCCCAAGGAGAGAGCTTTCCGCGCTGCGCGAGGC 471  
Db 66520 ATAGACGGCTTTCCTCCGCGACAGGCGCATAGAGAAAGGAGATGTGGCGCTGCGAGAG 66461  
Qy 472 CTCTCTC---CCGAGGCGGTGAGACCGGAGGTGTGAACAGGCGCTCATGAGCTCGG 528  
Db 66460 ATCTGCGCGGCGCGCGCGCGCATGCGCCCTATGTGAGAGGTGTGATGACCTGGGG 66401  
Qy 529 GCCACGGTCTGCTGCGCGGAAACGGCCCCCTTTGGCGGGCTTCCCTTGAAGGCTTCTG 588  
Db 66400 GCGACGGTGTGACAGCCCGCGGAAAGCGGCTGTGACCGCTGTCTGATGTGAGACGGCTCG 66341  
Qy 589 CGGAGGAGAGAGGCCCCCGGCGCTACCCCGCGCGCGAGAAAGCGCGGCGGAGAGAGAG 648  
Db 66340 GTGGCGGCGCGGAGGGCGCGACCGCGGAATCTGCCAGCGCGCGGAGAAAGGCTGTG 66281  
Qy 649 CGCTCTGCTGCGCTCTCTCTCTCGGCGGAGAGGGGTGCACTTGAAAGGCTTGAAGG 708  
Db 66280 CCATCCGCGACACCGGAGATGTGTGGCGCGCATGAAACAAAGGTATGTGCTTGAAGCG 66221  
Qy 709 CGCTCTCAG---GAGCTCTACGGCGGTCCCTCTTTCCCTGAGAGAGCTTCCCGGCGG 765  
Db 66220 CGGCGCGCGACCGGCATCTGGGCGGGCTGTCTTCTGCTCGGAGTTCCGACACACTTG 66161  
Qy 766 GAGCGGCGCTTCCGGGTGAGGTCTAGAGCCCTAGAGGAGGTGGCGACGCGCTCACCCAC 825  
Db 66160 GAATATGAAAGACTGTGCTGAGCGCGCGGAGATGCGACATGTGCTGCCGCTGGCGGAG 66101  
Qy 826 CGGAGGCTT 834  
Db 66100 CTGAGAGCAT 66092

RESULT 6  
AL646059 190050 bp DNA linear BCT 02-SEP-2002  
LOCUS Ralstonia solanacearum GM11000 chromosome, complete sequence;  
DEFINITION segment 3/19.  
ACCESSION AL646059 AL646052  
VERSION AL646059.1 GI:17427391  
KEYWORDS  
SOURCE Ralstonia solanacearum  
ORGANISM Ralstonia solanacearum  
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
Burkholderiaceae; Ralstonia.  
REFERENCE  
AUTHORS Salenoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S.,  
Arlat, M., Billault, A., Broctier, P., Camus, J. C., Catolico, L.,  
Chandler, M., Cholet, N., Claudel-Renard, C., Cunac, S., Demange, N.,  
Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, M., Schlex, T.,  
Siglier, P., Thebaud, P., Whalen, M., Winkler, P., Levy, M.,  
Weisenbach, J. and Boucher, C. A.  
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum  
JOURNAL Nature 415 (6871), 497-502 (2002)  
MEDLINE 21681879

PUBMED 11823852  
REFERENCE 2 (bases 1 to 190050)  
AUTHORS Boucher, C. A.  
TITLE Direct Submission  
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie  
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, INSERM  
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA  
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,  
Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,  
F31326 Castanet-Tolosan Cedex, Laboratoire de Génétique Cellulaire  
INRA, BP27, F31326 Castanet-Tolosan Cedex  
Christian.Boucher@toulouse.inra.fr  
http://sequence.toulouse.inra.fr/R.solanacearum.html.  
FEATURES  
source location/Qualifiers  
1..190050  
/organism="Ralstonia solanacearum"  
/mol\_type="genomic DNA"  
/strain="GM11000"  
/db\_xref="taxon:305"  
complement(163..750)  
/gene="RSC0382"  
/note="synonym: RS03350"  
complement(163..750)  
/gene="RSC0382"  
/function="miscellaneous; not classified regulator"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PROBABILE RESPONSE REGULATOR TRANSCRIPTION  
REGULATOR PROTEIN"  
/protein\_id="CAD13910.1"  
/db\_xref="GI:17427392"  
/db\_xref="GOA:O8Y2F4"  
/db\_xref="SPTREMBL:O8Y2F4"  
/translation="MAPTRYLVLRDPLDLEQVNLRLIRAGLAAGFVDVQVADLML  
PERIAQOPMVIVASESARDTIEHCATGAPRPYVLTFTNDPASRTIKAFANGI  
TATVDELKTRKAVLDVAAYRFEHRALRAFLDNAKTQLARKVLEBAKGLLMQOM  
QUSDEAFKRLRKAMDRIKIVAAQRYIDVAG"  
937..1638  
/gene="RSC0383"  
/note="synonym: RS03351"  
937..1638  
/gene="RSC0383"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAD13911.1"  
/db\_xref="GI:17427393"  
/db\_xref="GOA:O8Y2F3"  
/db\_xref="SPTREMBL:O8Y2F3"  
/translation="MTDFVATPPAVTGIFVPGAGEPFVRYVCGRNVAHAREMG  
DPREPPFPCKPADAVRVVADGTGQFVYPTETDCHYEILVAIGTGGDIVES  
AAQHIYGAIVIGDMTRDLQNAKKGKRWETGKAFDSGAPIGPIVPAQTIVSPRGA  
ITLSVNDKAHQHGLSDLIWSVETIAYLSRFLFELRPGLIETGPEGCPVSIIDLM  
VGRIDGVELHXYV"  
1642..2292  
/gene="RSC0384"  
/note="synonym: RS03352"



CDS  
1642..2292  
/gene="RSC0384"  
/EC\_number="5.2.1.2"  
/function="small molecule metabolism; degradation; amino acids"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PROBABLE MALEYLACETOACETATE ISOMERASE  
(GLUTATHIONE TRANSFERASE ZETA 1) PROTEIN"  
/protein\_id="CADI3912.1"  
/db\_xref="GI:17427394"  
/db\_xref="GOA:08Y2F2"  
/db\_xref="SPTREMBL:08Y2F2"  
/translation="MSIKLYNPPSSASFPRIATLEIKG.PYDYAPVHLKGEQAAD  
FVKNPBAIPVLCGTETVLNQSLAIVEYIEETHPEPTLPGSADRAHTRALALATA  
CEIHLNPNPVLKYLKHTFVNGDDARDWRHWRIGFPALETTLTQSPRTGACVDD  
TPTLADLCIVPOVFNKGRPDVAEDYPTLARVEHCMAOPAFORAPAAPDASAS"  
2289..2927  
/gene="RSC0385"  
/note="synonym: RS03353"  
2289..2927  
/gene="RSC0385"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PROBABLE TRANSMEMBRANE PROTEIN"  
/protein\_id="CADI3913.1"  
/db\_xref="GI:17427395"  
/db\_xref="GOA:08Y2F1"  
/db\_xref="SPTREMBL:08Y2F1"  
/translation="MTGPGPSAPLIGALSGLGASLIIVATGAONAYVLRQSLRE  
YVAGVUICACDMALIALSVAGKGTISHPAMLTRVRNAGAAFLLIYGRARRAN  
RGAERLQARNGDKASHAOVLASALATLPHVLTVDVLGALGGRYAMPANAPFAG  
GAMCASILMSLGLFGARLDEPVFARPVARVLDALIGAVMVAIALTLMLGG"  
complement (2924..3910)  
/gene="f1ay"  
/note="synonyms: RS03354, RSC0386"  
complement (2924..3910)  
/function="f1ay"  
/function="cell processes; cell division"  
/note="Product confidence : putative  
Gene name confidence : putative  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PUTATIVE CELL DIVISION PROTEIN"  
/protein\_id="CADI3914.1"  
/db\_xref="GI:17427396"  
/db\_xref="GOA:08Y2F0"  
/db\_xref="SPTREMBL:08Y2F0"  
/translation="MPDTPAPAPALADMOADDIETVTPPVVEORKKMGMSRLSGLSK  
TSSSLTLPGVGVDEALFEELRALLMADAGDATYTLDELARRKRAORIEYABEV  
KTLARDLILHLPLEKTMVGRDQPTVIMAGNGKGTISGLKHFOTGOSVLA  
LAAGDTFRAAREQLVWGRNNVTVAAGSGPAAVIFAVVNAARAGDIWADTA  
GLRPTQGLHMEELKRRVIRIGKMAATAPHELTLLVDNTGONALAQYKAFDDALGLTG  
LIVTLQGTAKGILAIARQRPVYFVIGVEVDLQPSAREFADALLG"  
4124..5629  
/gene="RSC0387"

CDS  
4124..5629  
/gene="RSC0387"  
/EC\_number="3.4.99.-"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PUTATIVE ZINC PROTEASE PROTEIN"  
/protein\_id="CADI3915.1"  
/db\_xref="GI:17427397"  
/db\_xref="GOA:08Y2E9"  
/db\_xref="SPTREMBL:08Y2E9"  
/translation="MRSEFPOSVSARTGRMAGITLVALLAGAGAPAPMAARQ  
AALITGAKGTQADPHRYLANGLRLIVEDHRAPTVAHWWVHAGSIDHNGTGV  
AHLIEMHFGITRAVGPGEFSRYVALGRENMTTDDFTMYFOIERSHADMALE  
ADRNANQLTIDKEFKPEBNVVKERRRIDSARATYEQMLAVFNAAPYRNPTIG  
PSDIDTMTVDADWDHMYKAPVNAATVAVGVDPDEVFROAORTYGLQPHALPRY  
AOBEPQGVYKRIWYKAPAEENPYVLAAYKAPLARDYKDVDPYALVLSAIDGYDNA  
RLPNLIVKGRDEKGRSLADDVNAAGYDGMNRPISFLIDGVADGHTTAETEOALRAOI  
RIARADGVTREALEIKRYVAQVVAIOIYRDSVFQGMETIGMAEMGTLSMRDLRILEKI  
KSVTPAQVOQVAKTYFTEDNLVVAITLPPFIDPKPKRPVGRREGGLR"  
5629..6972  
/gene="RSC0388"  
/note="synonym: RS03356"  
5629..6972  
/gene="RSC0388"  
/EC\_number="3.4.99.-"  
/function="miscellaneous; hypothetical/global homology"  
Query Match 14.8%; Score 144.2; DB 1; Length 190050;  
Best Local Similarity 57.0%; Pred. No. 7,4e-11;  
Matches 322; Conservative 0; Mismatches 213; Indels 30; Gaps 2;  
gene  
CDS  
36 CCGGGAACCGCCGCCCTCCCTGCGGGGGAGAGACCTTACCGCTCTGGT 95  
19001 CCGGCAACCGCCGCATACCTGCGCGTGGAGAACACCGGTGACCGTGGGCT 19060  
Qy 96 CTCGAGCTCTTTCGACGACGACCGCGGTGAGACAGCCCTCCCTATTACCCCGCTT 155  
19061 GTCGAGATCATGCTGACGACGACGAGGTGCGCGGTGCGCTACGACCGCGCTT 19120  
Qy 156 TCTGAGCGCTTTCACACCTGAAAGGCGCTGCGCGGCTTCCCTGGAAGAGTCTTAG 215  
19121 CGTGAGCGGTTTCCGACCTGACGAGCACTGGCCGCGACCGCGGACGACGATGATGCG 19180  
Qy 216 GGTCTGACGAGGCGGCGCTACTACCGCGGCGGGAACACTTCACCGCTGAGCCGGAAG 275  
19181 CGCCTGGGCGGCGCTGCGCTACTACCGCGGCGGGAACACTTCACCGCTGAGCCGGAAG 19240  
Qy 276 CGTGGAG-----GAGCTTCCCGGAGCTTGGCCGAGCTTCCGGGGGCTTCC 320  
19241 CGTGGTGGCGAGCATGCGCGGTGCTTCCGCGGATCCGAGCGCGTGGCGGCGTGGCC 19300  
Qy 321 TGGTCTGGGCGCTTACACCGCGGCGGCGGCGCTTCATGCGCTTCCGGGGAGCGGTTGGC 380  
19301 GGGCATGGGCGCTTCCACCGCGGCGGCACTGCGCGCTTTCGTAACGCGTGGCGGCGGC 19360  
Qy 381 GGCAGTGAACGAGACCTCCGAGGAGTCTCTCCCGCTTTCG-----C 425  
19361 CATCTCGAAGGAGACGTCACAGGCGGTGCGCGCTTTCGCGATCGACGCGCTTTC 19420  
Qy 426 CCGGGAAGCCCAAGAGAGAGCTTTTGGCCCTTGGCCAGGCGCTCTCCCGGAGG 485  
19421 CCGGGAACCGCGTCAAGAGAACGATGAGCGATGCGGAGACCGTGTGCGCCCGGCGC 19480  
Qy 486 CGTGAACCGGGGGTGGGAACGAGCGCTCATGAGCTGGGGGCAAGGTCTGCTGCC 545

Db	Accession	Definition	Accession	Definition
01	546	GAACGCGCCCTTGGCGGCGCTGC 570		
Db	19541	CGCGAGCCCGGCATGCGTACGCGC 19565		
RESULT 7				
3X640422				
LOCUS				
DEFINITION	BK640422	257701 bp	DNA	linear BCT 14-AUG-2003
ACCESSION	BK640422			
VERSION	BK640422.1			
KEYWORDS	complete genome.			
SOURCE	Bordetella pertussis Tohama I			
ORGANISM	Bordetella pertussis Tohama I			
REFERENCE	Alcaligenaceae; Bordetella.			
AUTHORS	1 Parkhill, J., Sebatina, M., Preston, A., Murphy, L. D., Thomson, N.,			
	Harris, D. E., Holden, M. T. G., Churcher, C. R., Bentley, S. D.,			
	Mungall, K. L., Cerdeno-Rodríguez, A. M., Temple, L., James, K., Harris, B.,			
	Quail, M. A., Achtman, M., Atkin, R., Baker, S., Basham, D., Bason, N.,			
	Cherevach, I., Chillingworth, T., Collins, M., Cronin, A., Davis, P.,			
	Doggart, J., Felwell, T., Goble, A., Hamlin, N., Hauser, H.,			
	Holroyd, S., Jagels, K., Leather, S., Moulé, S., Norbertczak, H.,			
	O'Neill, S., Ormond, D., Price, C., Rabinowitsch, E., Rutter, S.,			
	Sanders, M., Saunders, D., Seeger, K., Sharp, S., Simmonds, M.,			
	Skelton, J., Squares, R., Squares, S., Stevens, K., Unwin, L.,			
	Whitehead, S., Barrrell, B. G. and Maskell, D. J.			
	Comparative analysis of the genome sequences of Bordetella			
	pertussis, Bordetella parapertussis and Bordetella bronchiseptica			
	Online Publication			
	Nat. Genet. DOI:10.1038/NG1227-10.1038/NG1227			
JOURNAL	2 (bases 1 to 257701)			
REMARK				
AUTHORS	Sebatina, M.			
TITLE	Direct SubMISSION			
JOURNAL	Submitted (06-AUG-2003) Submitted on behalf of the Pathogen			
	Sequencing unit, Sanger Institute, Wellcome Trust Genome Campus,			
	Hinxton, Cambridge CB10 1SA, E-mail: ms@sanger.ac.uk			
FEATURES	Location/Qualifiers			
source	1..257701 "Bordetella pertussis Tohama I"			
	/organism="Bordetella pertussis Tohama I"			
	/mol_type="genomic DNA"			
	/strain="Tohama I"			
	/db_xref="taxon:257313"			
	76..1266			
	/gene="tuf"			
	/locus_tag="Bp611"			
	/note="synonym: tufA, tufB"			
	76..1266			
	/gene="tuf"			
	/locus_tag="Bp611"			
	/note="Similar to Escherichia coli elongation factor Tu			
	(EF-Tu) TufA or TufB (duplicate of tufA) SW:EFU BCO1			
	(P02990) (393 aa) fasta scores: E(): 1.1e-122, 81.47% id			
	in 394 aa, and to Burkholderia cepacia elongation factor			
	Tu Tu SW:EFU BURCE (P3167) (396 aa) fasta scores: E():			
	3.4e-139, 92.65% id in 395 aa. Identical to Bp0007,			
	100.00% identity (100.00% ungapped) in 396 aa overlap"			
	/codon_start=1			
	/transl_table=11			
	/product="elongation factor Tu"			
	/protein_id="CA643869.1"			
	/db_xref="GI:33564553"			
	/transl_table="1"			
	/locus_tag="Bp611"			
	/note="Similar to Escherichia coli elongation factor Tu			
	(EF-Tu) TufA or TufB (duplicate of tufA) SW:EFU BCO1			
	(P02990) (393 aa) fasta scores: E(): 1.1e-122, 81.47% id			
	in 394 aa, and to Burkholderia cepacia elongation factor			
	Tu Tu SW:EFU BURCE (P3167) (396 aa) fasta scores: E():			
	3.4e-139, 92.65% id in 395 aa. Identical to Bp0007,			
	100.00% identity (100.00% ungapped) in 396 aa overlap"			
	/codon_start=1			
	/transl_table=11			
	/product="elongation factor Tu"			
	/protein_id="CA643869.1"			
	/db_xref="GI:33564553"			
	/transl_table="1"			
	/locus_tag="Bp611"			
	/note="Similar to Escherichia coli elongation factor Tu			
	(EF-Tu) TufA or TufB (duplicate of tufA) SW:EFU BCO1			
	(P02990) (393 aa) fasta scores: E(): 1.1e-122, 81.47% id			
	in 394 aa, and to Burkholderia cepacia elongation factor			
	Tu Tu SW:EFU BURCE (P3167) (396 aa) fasta scores: E():			
	3.4e-139, 92.65% id in 395 aa. Identical to Bp0007,			
	100.00% identity (100.00%			

```

misc_feature
103..1203
/gene="tuf"
/locus_tag="BP3611"
/notes="HMPFfam hit to PF00009, Elongation factor Tu
family"
130..153
/gene="tuf"
/locus_tag="BP3611"
/notes="ScanRegExp hit to PS00017, ATP/GTP-binding site
motif A (p-loop)."
226..273
/gene="tuf"
/locus_tag="BP3611"
/notes="ScanRegExp hit to PS00301, GTP-binding elongation
factors signature. Confirmed by InterPro EMOTIF pattern
match."
1330..1641
/gene="rpsJ"
/locus_tag="BP3612"
/notes="Similar to Escherichia coli 30S ribosomal protein
S10 RpsJ SW:RS10 ECOLI (P02364) (103 aa) fasta scores:
E(): 8.4e-28, 76.477 id in 102 aa, and to Neisseria
meningitidis 30S ribosomal protein S10 NMA0133 TR:09JR21
(EMBL:AL162752) (103 aa) fasta scores: E(): 1.2e-29,
79.617 id in 103 aa"
/codon_start=1
/transl_table=1
/product="30S ribosomal protein S10"
/protein_id="CBE43870.1"
/db_xref="GI:33564554"
/translation="MKAKRIRLRKAPDYKIDQSAEIVDTAKTGAVVRPELPT
RIRRDVNLSPHVKTSRDPKPEIRTHORLMDIYDPDKYDVALMRDLPAGDVEIAL
Q"
1342..1629
/gene="rpsJ"
/locus_tag="BP3612"
/notes="HMPFfam hit to PF00338, Ribosomal protein
S10P/S20e"
1414..1461
/gene="rpsJ"
/locus_tag="BP3612"
/notes="ScanRegExp hit to PS00361, Ribosomal protein S10
signature. Confirmed by InterPro EMOTIF pattern match."
2044..2688
/gene="rplC"
/locus_tag="BP3613"
/notes="rplC"
2044..2688
/gene="rplC"
/locus_tag="BP3613"
/notes="rplC"
/locus_tag="BP3613"
/notes="Similar to Escherichia coli 50S ribosomal protein
L3 RplC SW:RL3 ECOLI (P02386) (209 aa) fasta scores: E():
8.6e-48, 67.777 id in 211 aa, and to Neisseria
meningitidis 50S ribosomal protein L3 NMA0129 TR:09JRH6
(EMBL:AL162752) (214 aa) fasta scores: E(): 2.4e-51,
69.487 id in 213 aa"
/codon_start=1
/transl_table=1
/product="50S ribosomal protein L3"
/protein_id="CBA3871.1"
/db_xref="GI:33564555"
/translation="MGRKGMRTIPFDESESTPVTLDVSNNRVTQKSLSDGYAAIIL
QVYTGRRARRVAKPQAGHYAKAGASGIIKEPRIDPARAAEPAAGAVIVAEVFPAA
GQGVADVGTSGKGFAGTIRKHNFGSQRASHGNSHSRVPSSIGQADPGIIPGKKM
SGHMGDVTRTVQNIQNDVVRDAERGLLMVGAVPAGAGGVIVRAVKAPAKKA"
2053..2652
/gene="rplC"
/locus_tag="BP3613"
/notes="HMPFfam hit to PF00297, Ribosomal protein L3"

```

```

Query Match      13.8%; Score 135; DB 1; Length 257701;
Best Local Similarity 60.2%; Pred. No. 1.1e-09;
Matches 250; Conservative 0; Mismatches 150; Indels 15; Gaps 1;

36 CCGGAGAAACGCCCGCCCTCCCTCGGCGGGGAGAGACCTTACCGCGTCTGAT 95
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 217833 CCGGACGCGCGCCACGATCTGCGCTGGGAGAAACCCGCGACCCCTATGCGATCTGGCT 217892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
96 CTCGAGGCTCTTTCGACGACGACCGGGTGAAGCAGCGCTTCCTATTACCGCGCTT 155

misc_feature      2341..2412
                /gene="rplC"
                /locus_tag="BP3613"
                /note="Scanned hit to PS00474, Ribosomal protein L3
signature. Confirmed by InterPro eMotif pattern match."
                2693..3310
                /gene="rplD"
                /locus_tag="BP3614"
                2693..3310
                /gene="rplD"
                /locus_tag="BP3614"
                /note="Similar to Escherichia coli 50S ribosomal protein
L4 RplD SW:RL4_ECOLI (P02388) (201 aa) fasta scores: E():
5.9e-41, 58.33% id in 204 aa, and to Neisseria
meningitidis 50S ribosomal protein L4 NMA0128 TR:Q9JBA2
(EMBL:AL162752) (206 aa) fasta scores: E(): 2.4e-51,
65.02% id in 203 aa"
                /codon_start=1
                /transl_table=11
                /product="50S ribosomal protein L4"
                /protein_id="CAE43872.1"
                /db_xref="GI:33564556"
                /translation="MDLKLNDQGNATPSADPTIFGRDFNEALVHQIVAFANARS
GNRAKDAEAVHGSHSTKKPKROKGTGRAGAMTSSPLMRGGRAFPMSPEENFSOKVK
KMVRGIRISLSIOLAREDRVAVDPFTLESPTKTLAAKLSKGLSDSVLITDNVDEN
VYLAIRNLPYVAVNEPRYADPLSLIHKKVLTIKPAIQLAEMLG"
                2729..3301
                /gene="rplD"
                /locus_tag="BP3614"
                /note="HMPfam hit to PF00573, Ribosomal protein L4/L1
family"
                3307..3603
                /gene="rplW"
                /locus_tag="BP3615"
                3307..3603
                /gene="rplW"
                /locus_tag="BP3615"
                /note="Similar to Escherichia coli 50S ribosomal protein
L23 RplW SW:RL23_ECOLI (P02424) (100 aa) fasta scores:
E(): 1.5e-14, 54.94% id in 91 aa, and to Neisseria
meningitidis 50S ribosomal protein L23 NMA0127 TR:Q9JX11
(EMBL:AL162752) (106 aa) fasta scores: E(): 6.9e-19,
63.83% id in 94 aa"
                /codon_start=1
                /transl_table=11
                /product="50S ribosomal protein L23"
                /protein_id="CAE43873.1"
                /db_xref="GI:33564557"
                /translation="WNAERLMQVILAPITTEKATFPAEKQOYAFRVADATPEIKA
AVELLFKVQVESVQVILNRGKVKRFGRFVGRRRNERKAYVALKDGEIDFAEVK"
                3325..3600
                /gene="rplW"
                /locus_tag="BP3615"
                /note="HMPfam hit to PF00276, Ribosomal protein L23"
                3604..4431
                /gene="rplB"
                /locus_tag="BP3616"
                3604..4431
                /gene="rplB"
                /locus_tag="BP3616"
                /note="Similar to Escherichia coli 50S ribosomal protein
L2 RplB SW:RL2_ECOLI (P02387) (272 aa) fasta scores: E():
2.4e-75, 67.03% id in 273 aa, and to Neisseria

```

Db	Accession	Definition	Source	Reference
Db	217953	CTCGAAMTATGCTGTGACGACGACGAGTGGCGACGATTCATTCGGTATTAACAGGGCTT	217955	
Qy	156	TCGTGAGCGCTTTCCACCTCTGAAAGCCCTGGCCGGGCTTCCCTGGAAGAGTCTTAG	215	
Db	217953	CTGTGAAAGCTTTTCCGATGTGTGGCCGCTTGGCGGCGCGCAGAGAAATGTATGCC	218012	
Qy	216	GGCTGGGAGGAGGGGGGGCTACTACCGGGGGGGAACACTCTCCACCG-----	263	
Db	218013	CTACTGGGACCGGCTGGGCTACTACGCGGCGCCCGCACTGTACACCGCTCGCGCAGGA	218072	
Qy	264	---CCTGGCCCGAAGCGTGTGAGAGACTTCCCGCAGACTTGTGCGGAGCTTCC	320	
Db	218073	AATATATGACGCGCTGCGCTGCGCGGCTTCCCGCGCGCGGAAGATGTGCCACCTTGC	218133	
Qy	321	TGCTGTGAGGCTTACACCGCGCGCGCGGTGCTTCATGCTTGTGGAGAGCGGTGC	380	
Db	218133	CGGACTCGGCGGCTCAGCGCGCGGCGCATCGCGCTTGTGCTTGTGCGAGCGAGCCC	218192	
Qy	381	GGCGGTGAGCGGAACTGTCCGAGGGGTCTCTCCCGCGCTTTCGCCCGGGAAGC	435	
Db	218193	CATCATGAGCGGCAACGTCAAGCGGGTGTTCACGCGGCACCTTCGGGATGAGGCG	218247	
RESULT 8				
LOCUS	BX640435	346259 bp	DNA	linear
DEFINITION	BX640435	Bordetella parapertussis strain 12822, complete genome; segment 13/14.		
ACCESSION	BX640435	BX470249		
KEYWORDS	BX640435	GI:33574489		
SOURCE		complete genome.		
ORGANISM		Bordetella parapertussis		
REFERENCE		Bordetella parapertussis		
AUTHORS		Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Alcaligenaceae; Bordetella.		
	1	Parkhill, J., Sebahia, M., Preston, A., Murphy, L. D., Thomson, N., Harris, D. E., Holden, M. T. G., Churcher, C. R., Bentley, S. D., Mungall, K. L., Cerdano-Tarrega, A. M., Temple, L., James, K., Harris, B., Quail, M. A., Achtman, M., Atkin, R., Baker, S., Basham, D., Baeson, N., Chervach, I., Chillingworth, T., Collins, M., Cronin, A., Davis, P., Doggett, J., Felwell, T., Goble, A., Hamlin, N., Hauser, H., Holtrop, S., Jagers, K., Leach, S., Moule, S., Norbertczak, H., O'Neill, S., Omond, D., Price, C., Rabinowitsch, E., Rutter, S., Sanders, M., Saunders, D., Seeger, K., Sharp, S., Simmonds, M., Skelton, J., Squares, R., Stevens, K., Unwin, L., Whitehead, S., Barrall, B. G. and Mankell, D. J.		
	2	Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica Online Publication Nat. Genet. DOI:10.1038/Ng1227-10.1038/Ng1227		
	3	(bases 1 to 346259) Sebahia, M.		
	4	Direct Submission Submitted (06-AUG-2003) Submitted on behalf of the Pathogen Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: ms@sanger.ac.uk		
	5	location/Qualifiers		
	6	1. 346259		
	7	organism="Bordetella parapertussis"		
	8	mol_type="genomic DNA"		
	9	strain="12822"		
	10	db_xref="taxon:519"		
	11	complement(293..937)		
	12	gene="ureg"		
	13	locus_tag="BPP3852"		
	14	complement(293..937)		
	15	gene="ureg"		
	16	locus_tag="BPP3852"		
	17	codon_start=1		
	18	/note="Ortholog of Bordetella pertussis (BX470248) BPP316"		
	19	/transl_table=11		

/product="urease accessory protein"  
/protein\_id="CAE39135.1"  
/db\_xref="GI:33574490"  
/translation="MHDISLTTTRTKTLPLRVGVGVSGKTTLLMEVCKAMYPQF  
DLIAITNDIVYKEDRLTLTGALPERILIGETGCPHTAREDAISINLIADOME  
QPPDDIVYESSGDNLAATFSPSLDLTLVITIVASGEKIPRGSGGIRKSDLIIN  
KTDLAPYAGADLAWEADTRMRDCKPFVWNLTKTGDLGVIAFLTEGLFRG"  
complement(509..892)  
/gene="ureg"  
/locus\_tag="BPP3852"  
/note="HMPfam hit to PF01495, HyB/Ureg  
nucleotide-binding domain, score 4.5e-73"  
complement(848..871)  
/gene="ureg"  
/locus\_tag="BPP3852"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)."  
complement(962..1624)  
/gene="uref"  
/locus\_tag="BPP3853"  
complement(962..1624)  
/gene="uref"  
/locus\_tag="BPP3853"  
/note="Ortholog of Bordetella pertussis (BX470248) BP3167"  
/codon\_start=1  
/transl\_table=1  
/product="urease accessory protein"  
/protein\_id="CAE39136.1"  
/db\_xref="GI:33574491"  
/translation="MHLSPPALPIGFSYQGLEAIEGLVHDEASTLAMESQLVT  
VMKAEAPLWCLFEARAGDDAAHGWNQWFPASSETRELROETOMRSIARLQOE  
LGWGAATRAVALRPAITLPAVHACAMMALPREAGAYVFWLENVAAIKGV  
PLQMGQRMLERLRAGLPVALDARARAGATPPRIDTFAPQVALSARHEQFSRLF  
RS"  
complement(1094..1546)  
/gene="uref"  
/locus\_tag="BPP3853"  
/note="HMPfam hit to PF01730, Uref, score 9.7e-74"  
complement(1647..2264)  
/gene="ureg"  
/locus\_tag="BPP3854"  
complement(1647..2264)  
/gene="ureg"  
/locus\_tag="BPP3854"  
/note="Ortholog of Bordetella pertussis (BX470248)  
BP3167A"  
/codon\_start=1  
/transl\_table=1  
/product="urease accessory protein"  
/protein\_id="CAE39137.1"  
/db\_xref="GI:33574492"  
/translation="MKIANTFKRGAAGGLDLSQAPGVTLTAERRRSRORLDDGR  
GELGVAIERGQTLRDGDVVAEDGVVVVRALEDAVATATPMQLARAVHIGNRH  
VLEIAEOHLOFEYDAVLIDMLAOLGGVTAHLEAVRFPDVGAVGGGRHGHDSFGD  
DYALAOAYAHAEAPHASHAGHGHVSHSGHCKGKHCHDAES"  
complement(1944..2249)  
/gene="ureg"  
/locus\_tag="BPP3854"  
/note="HMPfam hit to PF02814, Ureg urease accessory  
protein, score 1.7e-45"  
complement(2267..3982)  
/gene="urec"  
/locus\_tag="BPP3855"  
complement(2267..3982)  
/gene="urec"  
/locus\_tag="BPP3855"  
/EC\_number="3.5.1.5"  
/note="Ortholog of Bordetella pertussis (BX470248) BP3168"  
/codon\_start=1  
/transl\_table=1  
/product="urease alpha subunit"  
/protein\_id="CAE39138.1"  
/db\_xref="GI:33574493"  
/translation="WTRISRSAYAEIYGPVTVGVDGRVRLADTLTLLAEVEXHTTFC

misc\_feature  
complement(2273..3580)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="HMPfam hit to PF02802, Urease alpha-subunit,  
catalytic domain, score 0"  
complement(2969..3019)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="PS00145 Urease active site."  
complement(3548..3589)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="PS01120 Urease nickel ligands signature."  
complement(3584..3979)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="HMPfam hit to PF00449, Urease alpha-subunit,  
N-terminal domain, score 4.8e-83"  
complement(3979..4287)  
/gene="ureb"  
/locus\_tag="BPP3856"  
complement(3979..4287)  
/gene="ureb"  
/locus\_tag="BPP3856"  
/EC\_number="3.5.1.5"  
/note="Ortholog of Bordetella pertussis (BX470248) BP3169"  
/codon\_start=1  
/transl\_table=1  
/product="urease beta subunit"  
/protein\_id="CAE39139.1"  
/db\_xref="GI:33574494"  
/translation="MIPGEILTERGOIELVNGBPTLTIAVNEGDRPIOVGSYHRAE  
ANNAVLVDRRLATGYRLNIPAGNAVFEFGMRRTVLAVAGERRIFFGQKMGALK"  
complement(3985..4284)  
/gene="ureb"  
/locus\_tag="BPP3856"  
/note="This hit extended beyond the end of the feature by  
1 aa and was clipped.  
HMPfam hit to PF00699, Urease beta subunit, score  
3.2e-60"  
complement(4304..4921)  
/gene="urej"  
/locus\_tag="BPP3857"  
complement(4304..4921)  
/gene="urej"  
/locus\_tag="BPP3857"  
/note="Ortholog of Bordetella pertussis (BX470248) BP3170"  
/codon\_start=1  
/transl\_table=1  
/product="urease accessory protein"  
/protein\_id="CAE39140.1"  
/db\_xref="GI:33574495"  
/translation="MKAKSKRAMLGSSGAALMFSGAALAHPIGHILPGSMFAAGF  
WHPITGFDHLMLAVGMMSALTHTTARQAVLPMPLALLPAGMMGMAGVRLPAVE  
PYIWSLVLGLVLSRAVOGMAGFALPHGLAHGHELPSSGALFVAGFM  
LATIGHLAGLPAGPRLKHMNLMLSRALGVGIAGYALLFVGARV"  
complement(4844..4921)  
/gene="urej"  
/locus\_tag="BPP3857"  
/note="Signal peptide predicted for BPP3857 by SignalP 2.0  
HMM (Signal peptide probability 1.000) with cleavage site  
probability 1.000 between residues 26 and 27;

misc\_feature  
complement(2273..3580)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="HMPfam hit to PF02802, Urease alpha-subunit,  
catalytic domain, score 0"  
complement(2969..3019)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="PS00145 Urease active site."  
complement(3548..3589)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="PS01120 Urease nickel ligands signature."  
complement(3584..3979)  
/gene="urec"  
/locus\_tag="BPP3855"  
/note="HMPfam hit to PF00449, Urease alpha-subunit,  
N-terminal domain, score 4.8e-83"  
complement(3979..4287)  
/gene="ureb"  
/locus\_tag="BPP3856"  
complement(3979..4287)  
/gene="ureb"  
/locus\_tag="BPP3856"  
/EC\_number="3.5.1.5"  
/note="Ortholog of Bordetella pertussis (BX470248) BP3169"  
/codon\_start=1  
/transl\_table=1  
/product="urease beta subunit"  
/protein\_id="CAE39139.1"  
/db\_xref="GI:33574494"  
/translation="MIPGEILTERGOIELVNGBPTLTIAVNEGDRPIOVGSYHRAE  
ANNAVLVDRRLATGYRLNIPAGNAVFEFGMRRTVLAVAGERRIFFGQKMGALK"  
complement(3985..4284)  
/gene="ureb"  
/locus\_tag="BPP3856"  
/note="This hit extended beyond the end of the feature by  
1 aa and was clipped.  
HMPfam hit to PF00699, Urease beta subunit, score  
3.2e-60"  
complement(4304..4921)  
/gene="urej"  
/locus\_tag="BPP3857"  
complement(4304..4921)  
/gene="urej"  
/locus\_tag="BPP3857"  
/note="Ortholog of Bordetella pertussis (BX470248) BP3170"  
/codon\_start=1  
/transl\_table=1  
/product="urease accessory protein"  
/protein\_id="CAE39140.1"  
/db\_xref="GI:33574495"  
/translation="MKAKSKRAMLGSSGAALMFSGAALAHPIGHILPGSMFAAGF  
WHPITGFDHLMLAVGMMSALTHTTARQAVLPMPLALLPAGMMGMAGVRLPAVE  
PYIWSLVLGLVLSRAVOGMAGFALPHGLAHGHELPSSGALFVAGFM  
LATIGHLAGLPAGPRLKHMNLMLSRALGVGIAGYALLFVGARV"  
complement(4844..4921)  
/gene="urej"  
/locus\_tag="BPP3857"  
/note="Signal peptide predicted for BPP3857 by SignalP 2.0  
HMM (Signal peptide probability 1.000) with cleavage site  
probability 1.000 between residues 26 and 27;

misc\_feature

signal-peptide site"  
complement(join(4310..4375,4397..4462,4493..4549,  
4571..4621,4637..4702,4724..4789,4835..4900))  
/gene="urea"  
/locus\_tag="BPP3857"  
/note="77 probable transmembrane helices predicted for  
BPP3857 by TMHMM2.0 at aa 7-29, 44-66, 73-95, 100-117,

Query Match 13.8%; Score 135; DB 1; Length 346259;  
Best Local Similarity 60.2%; Pred. No. 16-09;  
Matches 250; Conservative 0; Mismatches 150; Indels 15; Gaps 1;

QY 36 CCGGAAAAACGCCGCCCTCTCCCTCGCGGGGGGAGAAAGACCTTACCGCTCTGCT 95  
Db 138937 CCGCCACGCGCGCCACGATCTCCCTGGCGAACAACCCGCACTCTATCGCATCTGGCT 1389366  
QY 96 CTCCGAGGCTCTTCGACAGACCCGGGTGAGAGCGCTCCCTATTCACCGCGCTT 155  
Db 138997 CTCCGAATCATCTCTGACAGACGAGATGGCGACGTCATTCGTTATTCAGAGGCTT 1390566  
QY 156 TCTGAGCGCTTTCACACCTTGAAGCCCTGCGCGGAGCTTCTTGAAGAGTCTTGA 215  
Db 139057 CCGTGAAGTTCCTCCGATGTGGCGGCTCTGGCGGGGGCGCGCAGAGATGTATGCG 1391166  
QY 216 GGTCTGCAAGGCGGCGGCTACTACCGCGCGGCGAACACTTCACG----- 263  
Db 139117 CTACTGCGCGGCTCTGGGCTACTACGCGCGCCCGCAACTGCACTGCGCTGCGCAGGA 1391766  
QY 264 ---CCTGAGCCGAAACCTGAGAGAGTTCCTCCGAGCTTTCGCGAGCTTTCGGGAGCTTCC 320  
Db 139177 AATCATGACAGCGCTCGGCGGCGCGCTTCCCGCGCGCGAAGATGCGACCTTGCC 1392366  
QY 321 TGGTTCGGGCTTACACCGCGCGCGGTGCGCTTCATCGCTTCCTGGGAGCGGCTGCG 380  
Db 139237 CGGCATCGGCGCGCTCAGCGCGCGCGCATCGCGCTTCGCTTACGCGAGCGCAGGCC 1392966  
QY 381 GCGCGTGACCGGAACGTCCGAGGGTCTCTTCGCGGCTTTCGCCCGGAAAGC 435  
Db 139297 CATCATGACGCGCAACGTCAACGGGGTGTTCACGCGCACTTCGCGATCGAGGGC 139351

RESULT 9  
BX640450  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
REFERENCES  
AUTHORS  
TITLE

BX640450 346287 bp DNA linear BCT 14-AUG-2003  
Bordetella bronchiseptica strain RB50, complete genome; segment 14/16  
BX640450 BX470250  
BX640450.1 GI:33577349  
complete genome.  
Bordetella bronchiseptica RB50  
Bordetella bronchiseptica RB50  
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Alcaligenaceae; Bordetella.  
1  
Parkhill,J., Sebahia,M., Preston,A., Murphy,L.D., Thomson,N., Harris,D.E., Holden,M.T.G., Churcher,C.R., Bentley,S.D., Mungall,K.L., Cerdono-Tarraga,A.M., Temple,L., James,K., Harris,B., Quail,M.A., Achtman,M., Atkin,R., Baker,S., Basham,D., Bacon,N., Cherevach,I., Chillingworth,T., Collins,M., Cronin,A., Davis,P., Doggett,J., Felwell,T., Goble,A., Hamlin,N., Hauser,H., Holtroyd,S., Jagsels,K., Leather,S., Moulé,S., Norderbach,H., O'Neill,S., Ormond,D., Price,C., Rabinowitsch,E., Rutter,S., Sanders,M., Saunders,D., Seeger,K., Sharp,S., Simmonds,M., Skelton,J., Squares,R., Squares,S., Stevens,K., Unwin,L., Whitehead,S., Barrrell,B.G. and Makhrelidze,D.J.  
Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica Online Publication  
Nat. Genet. DOI:10.1038/Ng1227-10.1038/Ng1227  
2 (bases 1 to 346287)  
Sebahia,M.  
Direct Submission  
Submitted (06-AUG-2003) Submitted on behalf of the Pathogen

Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA, E-mail: ms@eanger.ac.uk

Location/Qualifiers

1. 346287  
/organism="Bordetella bronchiseptica RB50"  
/mol\_type="genomic DNA"  
/strain="RB50"  
/db\_xref="taxon:257310"  
complement (65..1093)  
/locus\_tag="BB4251"  
complement (65..1093)  
/locus\_tag="BB4251"  
/note="ortholog of Bordetella pertussis (BX470248) BP0421"

/codon\_start=1  
/transl\_table=1  
/product="putative acyl-CoA dehydrogenase"  
/protein\_id="CAE34615.1"  
/db\_xref="GI:33577350"  
/translation="MSTVMODLVYTPFAPALAGLCPELYRELEAAADPAGAARVNN  
EVDDLTGDAIISPARGAGSLSDAEGILRAAGNMLPPFPADTTLARLLRLAAGQA  
VPDPIALGVALLPHGAWRCAPLAGVALVAOVAVEHGDIATLPAEPADVQGRPRPA  
SGAPVRGTGPAGRAGAPGAIVQAMNMNADAGMACAMQAVLDRCVAVYQERQFRPA  
ALGFQAVOODISMLAEQVAIVAMAARLACASGLFPDPALAICARLRACEAVPVCA  
LAHVHGAIGITEELPGVTARLHERNRATGMSEBDCAELLGRRVLADGRITLDVVR  
GMSAATA"

complement (128..565)  
/locus\_tag="BB4251"  
/note="HMMPfam hit to PF00441, DE Acyl-CoA dehydrogenase,  
C-terminal domain, score 6.3e-06"  
complement (743..793)  
/locus\_tag="BB4251"  
/note="PS00372 PTS BIIA domains phosphorylation site  
signature 2."  
complement (1120..2292)  
/locus\_tag="BB4252"  
complement (1120..2292)  
/locus\_tag="BB4252"  
/note="ortholog of Bordetella pertussis (BX470248) BP0420"

/codon\_start=1  
/transl\_table=1  
/product="probable acyl-CoA dehydrogenase"  
/protein\_id="CAE34616.1"  
/db\_xref="GI:33577351"  
/translation="MGAILDTITLTAQLAIIPAHDELRPAVARFLDEALEKEMPADRR  
ARSQMGDIAEFSSALAQCGWLGILTLPVEYGAGSINRARVLSBELLAAPAVSANHI  
ADQMTAILRFSGPAORAFYLRIIRGEAFPAGMEPTGDSAPRTAPVAGAPVAPSAHI  
WLNCRKITWTNHRSHYCALVRTSGAPEDRRGCSOMFDALPEIETRPIDINGQ  
DAHCEVLPFNVLPHDALYGBSGSCROYMAELERSSPERTYSMTLDDGLATL  
RDDAPRAOVCLAGRALARLARLAVRSNSLVAQSLEGGASDLAAVLVKLGTFFEQA  
VPEILIGOLHASQPAPDDVDIMRTLAVTLINPFISLFGSTRHLIRIELGLR"  
complement (1579..1884)  
/locus\_tag="BB4252"  
/note="HMMPfam hit to PF02770, DE Acyl-CoA dehydrogenase,  
middle domain, score 4.9e-11"  
complement (1756..1806)  
/locus\_tag="BB4252"  
/note="PS00237 G-protein coupled receptors signature."  
complement (1888..2244)  
/locus\_tag="BB4252"  
/note="HMMPfam hit to PF02771, DE Acyl-CoA dehydrogenase,  
N-terminal domain, score 0.00021"  
2418..3395  
/locus\_tag="BB4253"  
2418..\_3395  
/locus\_tag="BB4253"  
/note="Ortholog of Bordetella pertussis (BX470248) BP0419"

/codon\_start=1  
/transl\_table=1  
/product="Layer-family transcriptional regulator"  
/protein\_id="CAE34617.1"  
/db\_xref="GI:33577352"  
/translation="WMTGQTVIHVIASPAHHWRRAVELROUKOLLVSETLNFHRAAR  
LHAAGPLSTAIRKLKEGLEGVOLFERRPALQTPRAAEVVLRARALTFVADIEQRPA

OEASGKQMKXGVFGSSTYSLSMPLSLAERKDYPRVDIVIBESTVDILRLDLNHT  
LDVALYRFPVLEPSTARIITLQDHLMLAVPASRPAQRDVALDELADPEFGYSRT  
HVPQMALIMYAFQYGVPHIAQEALQVQTIISLVESGLAIVPKVACRQSGSVR  
LVNVPQALAEITIKVGIALAVHPDNATPTTANFVMACRLMOTEPBAQGAR"  
2484. .2663  
/locus\_tag="BB4253"  
/note="HMMpfam hit to PF00126, DE Bacterial regulatory  
helix-turn-helix protein, lyxr family, score 7.5e-23"  
2526. .2618  
/locus\_tag="BB4253"  
/note="PS00044 Bacterial regulatory proteins, lyxr family  
signature."  
2733. .3356  
/locus\_tag="BB4253"  
/note="HMMpfam hit to PF03466, DE Lyxr substrate binding  
domain, score 3.1e-43"  
complement (3406. .4377)  
/locus\_tag="BB4254"  
complement (3406. .4377)  
/locus\_tag="BB4254"  
/note="ortholog of Bordetella pertussis (BX470248) BP0418"  
/codon\_start=1  
/transl\_table=11  
/product="putative exported protein"  
/protein\_id="CAB34618.1"  
/db\_xref="GI:33577353"  
/translation="MHSRLTALALSVALLAAGSGARASDYPAPKIVLSPGSG  
ADTARFLSKHLAERPKRPVNVNRCGSGFTAAQAVATAPDGTTLFVASPMPTN  
AAVFKLPYDPAVCDPAPVAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP  
TATQVLELFEHQNGIKATHPYKGTSAADVAGVQVSVADVASALPIRGKLG  
RPLAVSTRIRIKDLPVPTMOESGNKGFAYMTAFPAKIDPAIVASDVAVALV  
RSQGRKAFMEQIGSEPFVCGPDLAFCDEDELASMRIRAKLANIQEE"  
complement (4306. .4377)  
/locus\_tag="BB4254"  
/note="Signal peptide predicted for BB4254 by SignalP 2.0  
HMM (signal peptide probability 1.000) with cleavage site  
probability 0.873 between residues 24 and 25;  
signal-peptide site"  
complement (3418. .4236)  
/locus\_tag="BB4254"  
/note="HMMpfam hit to PF03401, DE Uncharacterized protein  
family, UPF0065, score 4.4e-85"  
complement (4444. .5229)  
/locus\_tag="BB4255"  
complement (4444. .5229)  
/locus\_tag="BB4255"  
/note="ortholog of Bordetella pertussis (BX470248) BP0417"  
/codon\_start=1  
/transl\_table=11  
/product="probable enoyl-CoA hydratase"  
/protein\_id="CAB34619.1"  
/db\_xref="GI:33577354"  
/translation="MTDFLRHRRHGVSLLTMDPRRTNALSDADVAVALVQACARIG  
ADLSRAVLTGSHGVSFGNLTSTVGAGLSEPHSROATRSGLQRIPLVILN  
EVPITAVNGPAIGACDCLRLARQASFNESVVKGLTPGDGAWLLRIVYG  
BTRACELAFTRGSDIAPARRELIVSEVFDALLPALALAOEIAQSHGHALRTRK  
ILRSEQHRLDLTLEISAGFCFLAHTOEHGALDAHLSRRKG"  
complement (4660. .5190)  
/locus\_tag="BB4255"  
/note="HMMpfam hit to PF00378, DE Enoyl-CoA  
hydratase/isomerase family, score 3.9e-50"  
complement (5222. .6454)  
/locus\_tag="BB4256"  
complement (5222. .6454)  
/locus\_tag="BB4256"  
/note="ortholog of Bordetella pertussis (BX470248) BP0416"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CAB34620.1"  
/db\_xref="GI:33577355"  
/translation="MSGPLSEGVRLIDVTVMGPYATQILGDLGADVIVKVEPPAGDNV

RGIGRHHADMGIFLHANSKRSIALDLKNPAGREALKYAATCDLVVNVPRQAMA  
RLGISVAVAOANPSILVGMVGYDORPVARAAYDILGAGISLSVAGSDIP  
RYAPSAWDRIIVGISAAVAATAGLYNPARTOGOGATVPMETMAHILIGHMGOTF  
EPPRIGPVPYRILNADRPVATRRGHTCVLLYADKHKAPFELGQGRDADRFRD  
IGTRQNHAIYRLVSDIASRTTBEMLAETADITAMPVHTVASTLDRPHLEATIG  
FGMDHSEGAIRMAVFSRWSGTRPPQGRHAPRLGHSVDVLEKAGYDDEALVALR  
NGAALIAARQLPEGTDDH"  
complement (5681. .6250)  
/locus\_tag="BB4256"  
/note="HMMpfam hit to PF02515, DE CAIB/BAIF family, score  
3.6e-24"  
complement (6650. .7405)  
/locus\_tag="BB4257"  
complement (6650. .7405)  
/locus\_tag="BB4257"  
/note="ortholog of Bordetella pertussis (BX470248) BP0415"  
/codon\_start=1  
/transl\_table=11  
/product="putative membrane protein"  
/protein\_id="CAB34621.1"  
/db\_xref="GI:33577356"  
/translation="MDYVSLAQQPAPMAALATLIMEVVLGIDNLIIFISILTNKLP  
OHQTKARRIGSMALIRLALGLTAVITVLTQVPELFGHGSMDVILVAGGLFLV  
WKATKEIHNVDRPQDMFGKAAITLGFAPAAVAQILMDLVESIDSIITAVGKTEHI

Query Match 13.88; Score 135; DB 1; Length 346287;  
Best Local Similarity 60.28; Pred. No. 1e-09;  
Matches 250; Conservative 0; Mismatches 150; Indels 15; Gaps 1;  
QY 36 CCGGGAAGTCTTTCGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 155  
219781 CCGGCAAGTCTTTCGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 219840  
QY 96 CTCGAGAGTCTTTCGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 155  
219841 CTCGGAATCATCTGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 219900  
QY 156 TCTGAGAGTCTTTCGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 215  
219901 CTTGGAAGTCTTTCGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 219960  
QY 216 GGTCTGAGAGTCTTTCGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 263  
219961 CTACTGAGAGTCTTTCGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 220020  
QY 264 ---CTTGGCCGAGAGTCTTTCGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 320  
220021 AATCATGAGAGTCTTTCGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 220080  
QY 321 TGTCTGAGAGTCTTTCGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 380  
220081 CGGATGAGAGTCTTTCGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 220140  
QY 381 GCGAGTGAAGAGTCTTTCGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 435  
220141 CATCATGAGAGTCTTTCGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 220195  
Db 220141 CATCATGAGAGTCTTTCGACAGACCCGGGTGAGCAGCCCTCCATTACCGCGCTT 220195  
RESULT 10  
SC0939116 293050 bp DNA linear BCT 11-FEB-2003  
LOCUS SC0939116  
DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 13/29.  
ACCESSION AL039116 AL035636 AL049573 AL049628 AL049645 AL049661  
AL049707 AL049763 AL049819 AL049841 AL079345 AL133236 AL133252  
AL645882  
VERSION AL039116.1 GI:24413781  
KEYWORDS Streptomyces coelicolor A3(2)  
SOURCE Streptomyces coelicolor A3(2)  
ORGANISM Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;  
Streptomycinae; Streptomycetaceae; Streptomyces.  
REFERENCE Bentley, S.D., Chaber, K.F., Cerdeno-Tarraga, A.M., Challis, G.L.,

Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kleser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. M., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C. H., Kleser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S., Rabinowitsch, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorek, A., Woodward, J., Barrett, B. G., Parkhill, J., and Hopwood, D. A.  
Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2)  
Nature 417 (6885), 141-147 (2002)

JOURNAL MEDLINE 21996410  
PUBMED 12000953  
REFERENCE 2 (bases 1 to 293050)  
AUTHORS Bentley, S. D.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces Sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA. E-mail: sdb@sanger.ac.uk  
COMMENT On or before Oct 26, 2002. This sequence version replaced g1:4539555, g1:4582361, g1:4585581, g1:4585828, g1:4678647, g1:4678625, g1:4678897, g1:457084, g1:4803679, g1:4808332, g1:5123647, g1:6469451, g1:6491805.  
FEATURES  
source location/Qualifiers  
1..293050  
/organism="Streptomyces coelicolor A3(2)"  
/mol\_type="genomic DNA"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
57..713  
/gene="SC03282"  
/note="synonym: SCE39.32"  
57..713  
/gene="SC03282"  
/note="SC03282", possible precorrin-8X methylmutase, len: 218 aa; similar to many e.g. SW:COBH\_PSEDE Precorrin-8X methylmutase from *Pseudomonas denitrificans* (210 aa) fasta scores; opt: 868, z-score: 1000.1, E(): 0, (64.2% identity in 204 aa overlap) and SW:COBH\_MYCTU Precorrin-8X methylmutase from *Mycobacterium tuberculosis* (208 aa) fasta scores; opt: 852, z-score: 982.0, E(): 0, (64.1% identity in 206 aa overlap)."  
/codon\_start=1  
/transl\_table=1  
/product="putative precorrin-8X methylmutase"  
/protein\_id="CAB40340.1"  
/db\_xref="GI:4582393"  
/db\_xref="GOA:O9X8F5"  
/db\_xref="SPTREMBL:O9X8F5"  
/translation="MTPTDSDKNYEYEEKDGAIIYQSFATITPAEDLSGLPADVGR VAVMIMHACGVNVDLVRDLSWSPGVASRARELRAGAPVLCVRNVASGVTRARLPADN DVCTLSDPAPVELARLIGTRSAALRLMRDLRDCAVVAAGNAPTLFRLEMTIGGS APRPAIVIGVGVGAESKEALAAHPSGVEHLVVRGRGSGALAAAVAVIAISEES"

gene complement(745..1512)  
/gene="SC03283"  
/note="synonym: SCE39.33c"  
complement(745..1512)  
/gene="SC03283"  
/note="SCE39.33c, possible precorrin-6X reductase, len: 235 aa; similar to many e.g. SW:COBK\_RHOSO Precorrin-6X reductase from *Rhodococcus* sp. (248 aa) fasta scores; opt: 873, z-score: 978.5, E(): 0, (55.1% identity in 247 aa overlap) and SW:COBK\_MYCTU Precorrin-6X reductase from *Mycobacterium tuberculosis* (244 aa) fasta scores; opt: 690, z-score: 775.5, E(): 0, (50.4% identity in 244 aa overlap)."  
/codon\_start=1  
/transl\_table=1  
/product="putative precorrin-6X reductase"  
/protein\_id="CAB40341.1"  
/db\_xref="GI:4582394"  
/db\_xref="GOA:O9X8F6"

/db\_xref="SPTREMBL:O9X8F6"  
/translation="MHVLLIGTTEARRLELAAGOPAGRVYTTSLAGRVAPRTP GEVRVGFSGAGELGALMLREHAGVLLVDATPHPECTTRNAHAAATAGVLLVLRP GMRVVEDVHNEGSLAEALALPALARVFLTTGRGLAVFADLPMPVLYRVADAP EGMRPTEVLRVGPPTLGGERLLRRHIDIVATVDSGAATAPKLTAREAGLPV VVVRPVPVEGVVAVRPGQAAVAVARARSPRRP"  
1683..3818  
/gene="SC03284"  
/note="synonyms: SCE15.01, SCE39.34"  
1683..3818  
/gene="SC03284"  
/note="SCE15.01, partial CDS, conserved hypothetical protein, len: >436aa; similar to C-terminal region of TR:CAA22732 (EMBL:AL035161) another conserved hypothetical protein from *Streptomyces coelicolor* (860 aa) fasta scores; opt: 922, z-score: 1011.5, E(): 0, (39.5% identity in 425 aa overlap)."  
SCE39.34, unknown, len: >307aa; similar to TR:E1370406 (EMBL:AL035161) conserved hypothetical protein from *Streptomyces coelicolor* (860 aa) fasta scores; opt: 254, z-score: 277.0, E(): 4e-08, (29.2% identity in 277 aa overlap) TR:086875 (EMBL:U02894) hypothetical protein from *Streptomyces lividans* (547 aa) fasta scores; opt: 226, z-score: 249.7, E(): 1.3e-06, (33.0% identity in 197 aa overlap)."  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="CAD5531.1"  
/db\_xref="GI:24413782"  
/db\_xref="GOA:O8CJX1"  
/db\_xref="SPTREMBL:O8CJX1"  
/translation="MKRARGAKAGGAATGDPITGDAAYLSALFANSRQGLFVDSQ KVRKYNPAGRGVRLPREEDVVGQIAFVFGFVEVEIGLIDEVLAGEALRGVLRG RABEGPRTTAAVEVSLPRLDGDGKGLGVGVEDYERQAAORLAVLSRVHTVGS TLVDRTGDELVALVAFADATVLDLDGVAABAGPPEVPLRVAFAPPA EVSRKEGSDRPLPEPTPYTOVNMARLAVRSDDPMLSDPALBPLVLRVGHSMI VAPILADIVYGLITLVRHRTDPREEDLVARQAATASVHLMNSVYREHVVAA LQRLDGLAIPLUSAVETAAHYLPESAGGMEFVDVPLASGRVALVVDVGHGLEAA LTMGLRLAVTLRLDDELTDLHDEVASQLDASGPGSNTVATCAVAVPVR HCTMVRAGHAPVVIDGVPRAVDVQGGPLDGGVHVFPAVLEAPGLLHLTN GTLRGNDGEBAARRLEHILASTRTLOELCTDGGVHMAFSRHDVALLRTRALP PEHVAWMTLPADASVGTARLVDRGLASGLDAAVTTGLVSELVTNARIYKGPV RLRLIRGRGLTFTVDANSASPHLRARNDGCRGLFIYMRSLTHGVHSHDKT IWSQRJRDAGPDTSGLDAPFVTDAAGL"  
complement(3855..9173)  
/gene="SC03285"  
/note="synonym: SCE15.02c"  
complement(3855..9173)  
/gene="SC03285"  
/note="SCE15.02c, large glycine/alanine rich protein, len: 1772aa"  
/codon\_start=1  
/transl\_table=1  
/product="putative large glycine/alanine rich protein"  
/protein\_id="CAB41271.1"  
/db\_xref="GI:4678899"  
/db\_xref="SPTREMBL:O9X877"  
/translation="MATAPKAAKTGHTDGTGADAGAYEVLRDRLAQTADLARRAGL LMARVEPSSARLEASTERTREHGVPCDDLAAGDALLSGTRGRTDAVA DVTLYRDNLDPESAVPGLDPPARFVRPDRVAYRRORLRLRBPVGRRLAVPR TGEVADIRVRLWELTDDGRAFLDANGERDVPPSHDFWTAATEDHYLGHNPV SVRKEFFVSAVAGVLLVTKLDDRTAGGVSEPVDEPQSLADDIARLGAALILR VRPEETDRLVNTLTKSVRLDGIACRRLPDDGIYFPGGYCLATVGHVVEL DAAGLFEFERVSPNGEDVLYAFHARSESGLLSYTTIRKEVNPPLPCRMALAEIG TPTVLRDGEPAQVHVPVQWHSPPYVDTHAAAPASGSPARVGNMDLVGICASF VAGAVDGGITTAISGRALAAACVRAAGHMLGADGLDGLAAVARETBOYLAER ETYVDRRAAEARDEAERIASVYRRLRGEARPEMAAAYRGLTELHRAHETLTYVE MRYVADGIDALGEAEASIAEIGRRAVAFARBDADVDAQADVLEALVADAEATVA EAGVVARLELDLGLRTVDDVAELDVGATVATVRLVAAVLGGVNRARATLDR RRLALDREGRAAEFTAPALGOAVTALAALADPERDCDQRLARLALAELEDFEAFR DGFALAEADRTEITVDAARKQALSTGRARBARQALRIWETIARCATLADAD AVSTYFASDMPAKVRRTAEBRLALGDSVAAEELDGLKSKARQASALNRRTLVAD







gene  
complement (574. .1056)  
/gene="m117508"  
CDS  
complement (574. .1056)  
/gene="m117508"  
/codon\_start=1  
/transl\_table=11  
/product="peroxiredoxin-like protein"  
/protein\_id="BAB53954.1"  
/db\_xref="GI:14027000"  
/translation="MTISVGDKLPEATFKMTADGAKATSAEIPGKVVLPFGVGA  
FTPTCSNNHLPGYLENDAILARGVDITAVSVVDVHVMGAMARFTGSEDKIILLAG  
SADPAKAVGLDNDSAGMGMLRSKFSMIVDDGKVTALNITETKGVDESGAAHILGOL  
"  
gene  
complement (1144. .1950)  
/gene="m117509"  
CDS  
complement (1144. .1950)  
/gene="m117509"  
/note="unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB53955.1"  
/db\_xref="GI:14027001"  
/translation="MOSIKTILLGAAIGSASVLPASASSSAMVNSEGGKVLVTSQKP  
DEAGRIQGVLDIALKPGMKTWRDPCGAGVPOLNVSATNIADATLSFPAPORHDDG  
VKGAGVNVPSLPIVFTLSTAKDSAVIDADVETLGCITTCIPVOTRLTVDPSDDPN  
ADDAALYKASFPALPAPARSDPGINVLPGDHETLVYASSPFGDEADDFPAGSERDYM  
FGAPRSREKDGKILFTVPIIDRPSATPTDGLHTLTLSAGAVEGLLFP"  
2122. .2730  
/gene="m1r7511"  
2122. .2730  
/gene="m1r7511"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB53956.1"  
/db\_xref="GI:14027002"  
/translation="MDLIRHKXTAAGRGFLDDOFLIAMPKMKDRFTRSVLYICASHD  
EGAMGLINOTQOMLPFDLLVQLGIMNBOEIRIPAPQRPVVRNGPVRDSRPVUH  
SGDRVSSSLVSDICLTATVDILRAISSGRGRHALMALGYSWGAGOLETIEAN  
GWLTPASPELFPADIERYDRILASIGIDLALSLAAGA"  
complement (2752. .4416)  
/gene="m117513"  
complement (2752. .4416)  
/gene="m117513"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB53957.1"  
/db\_xref="GI:14027003"  
/translation="MAGSGDTWMDVLRDRVYTKPVSITQLGAPNSLGGAAHMLP  
VLHADDDTFRITLDVLEHRRGRVAQNFRLRADGAYHWFALARPVIGSDGEVIRC  
VTAVDYTEQKSEERLHDVHNDLITGLPRELFPNRLAIISARTEEKVRFVIV  
IDIDRFKQVNDGIGISAGDTILITARLRLHLSKXSLSPADGOFALMLSEODPA  
RIGVADAITHAINNPITPAKREIVLTAIGLITWTSASAEEDMVADELAMHQAAR  
FGGDRIEPRPARFTVCTDRLOPESDILRAIIRERETLAIOPYITRLDGSVAGEBAL  
RWDPRRGMIPPADFIPVASCGLIVOLGFPAMQAAEDLAGOKQIGDAPLSVNL  
SSROLIRDLVSDVRSVIARANKLPKCFRLTESLVMNDEQTAHVLTKLKQIGIGL  
SLDFTGYSGLATLRFPFDTIKIDSFYVDNTPKRAVLKSNVMAHEIGLSVAE  
GISDERDALELRMGCEVOSPFMGAMPDQVVLKTLKEGYPLTQA"  
complement (5644. .6240)  
/gene="m117514"  
complement (5644. .6240)  
/gene="m117514"  
/codon\_start=1  
/transl\_table=11  
/product="ribosomal-protein-alanine N-acetyltransferase"  
/protein\_id="BAB53958.1"  
/db\_xref="GI:14027004"  
/translation="MPALPFFRRDLPALKGNLVTLRVPYTNDRYEMSTLRGSEAPLE  
PWEPRMTPELDRTPAMRLRISRYREDDYAQGTALAFIFEEKSGCLAGITLGNIRHV  
SOGCHVWPIGDERGRGLMTDAKVAVFAFDTLRRLHRTIEACTIPONHSIRVLEKA  
GRREGILRSYLRINGIMODHYLVARLADPELDAGTQD"  
complement (6256. .7548)  
gene

CDS  
/gene="m117516"  
complement (6256. .7548)  
/gene="m117516"  
/codon\_start=1  
/transl\_table=11  
/product="processing proteinaase"  
/protein\_id="BAB53959.1"  
/db\_xref="GI:14027005"  
/translation="MGYVSRSLNSGLTVAETLPSITSVALGAMVYKSGARBERBEG  
MAHLEHMAFGTKRTAFETASIEIDVGEINATISVETTSYARVLSDVPLAVDI  
LADLQSEEDPOLEBEROHVILQELGAADTPDIDVDRFTETAFPHQTGRSILGT  
PDVYKSTFSKQHDPIEROYGAEHVVVAAGDIXHDFNREVERKOLGFSRKAOSTIP  
QYAOYVGDFREDRDLMDAQIVLGFEGRAVHRDFVAYOVLSMILGGMSRRLFOEVR  
EKRGICYSYAFHMGFSDGTGIFGHATGOSDIAELIYVPIIDELQKXENITLOEILR  
ARAQYRAGLIMSASPSARASQIAROLLGRLPAKELMERLSALTVERLTDSSKM  
FSTKPYLTAVGPVGLTAPYALIDSLSGTQTLARLAV"  
complement (7563. .8960)  
/gene="m117517"  
complement (7563. .8960)  
/gene="m117517"  
/codon\_start=1  
/transl\_table=11  
/product="threonine synthase"  
/protein\_id="BAB53960.1"  
/db\_xref="GI:14027006"  
/translation="MQYVSTRGEAPALGFSDAVIAGIARDGCLVYPREMPOFSSEIR  
AMGLAYPDIAIRVLSFFLGSEIPAFVFEVLVREAYATFFHEAYCPVLVOTGNTFVLE  
LFHGPTLAFKDVAMQTLARLMDHVARDDHATTVGATSGDTGGAALDAFRAGSRTOI  
FILEPHGRVSPVOQRMTTSAENVHALATEGNDDCGLLKMFNDHAFDRVALSG  
VNSINMARIMAOIIVYFSSALSIGAPRPVSPFTVPTNGFDIFAGYAAKKGLPIERL  
VIATNDNDILARTPATGEVYRTKGFATPSTMOIVSNEFRLLFEASNDRAATVVR  
MDGLKSGAFITSEIREINGRSEFDDARATMDVAAITRSTLASNYLDPHTAAAH  
VAGKASGAVPMVVLGTAFHPAKFPAAVEAASVSPALPWLGLMTTEBKTYLPSDL  
KMVEDYVSRRAAR"  
8930. .9817  
/gene="m1r7518"  
8930. .9817  
/gene="m1r7518"  
/note="unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB53961.1"  
/db\_xref="GI:14027007"  
/translation="MLPRATSHIACLSPVAAWLSVHHKVNFCAGSSRIFAPALITYT  
SFTRGKCSFMAFRITLSGRFLAGALTGFMILAAAGCSGSDGILNLFGEKDPTRAPPP  
QDPRIASQULNAYCPRVTYVDGTAFETVANEVKARPKLKTGDAADAADAQAADA  
AAATPPDSSARIIVYQASIADVTRDCSADQLAMKIVAGKIYGPFTPTGTTMPYR  
TAVMGHDVLYSQIHQYQVOTDPSVATQVFFTDSNVVPAFTADQYVAGYDETAP  
KATTDKPKTKHRRKAATTN"  
complement (9860. .10549)  
/gene="m117519"  
complement (9860. .10549)  
/gene="m117519"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB53962.1"  
/db\_xref="GI:14027008"  
/translation="MPQPDVIVFDCDVLVDSBIIAARVAEELITLAGYSIAEIAE  
SVAGLTKFDILMRIEESKILPFOASLIDRAEBLVDVRAIEGVREAVSVTQ  
RCICNSGRSRIIRIFMLEKXILLPPPARIRISALEIPEKTKRPADVDLPAERBAGAP  
ANTFVIEDSVGVAGARBAAGRVIGFTGASHSTPGHADALTGCAETVIRRMALKSY  
IALSEWSADA"  
10726. .11859  
/gene="m1r7520"  
10726. .11859  
/gene="m1r7520"  
/codon\_start=1  
/transl\_table=11  
/product="adenine DNA methyltransferase"  
/protein\_id="BAB53963.1"  
/db\_xref="GI:14027009"  
/translation="MSAVRLHLSHAPOOSEWLDITLKQDCVAAALRLPEKSVDVIF

ADPPYNLQJLGDHLPDQSKVDADVHDWQFENFEAYATFTRAWLLAARVUKENGT  
WIGSYHNI FRVAGAKODLGFWILNDVYWRKTPNPNRGRFQAHETMTWASRDQK  
GKGYTNEYALKASNDLIQMSDMLFPICTGSERLKNDYDIDANRIRIDAVPLEDADL  
ASTKXDI VLDPFPGSGTGAVKRIGFVIEREQYIDANRIRIDAVPLEDADL  
TYLTKRABRPVAFSLIDTGLMTBGVTLYDKRMWAKVRADGVVALGDSAGSIHKI

Query Match 13.4%; Score 130.2; DB 1; Length 346547;  
Best Local Similarity 55.8%; Pred. No. 4.5e-09;  
Matches 324; Conservative 0; Mismatches 233; Indels 24; Gaps 3;

33 GATACGGGAAAAAGCCCCCCCCCTCCCTGGGGGGGGGAGAGACCCCTTACCGGCTCT 92  
13560 GTGGGGGGGAGACCCCGGAGCATGACGCGCGGCGCCGACCCCTATCGACTTG 13501  
93 GGTCTCCGAGGTCTCTTCTGACAGACCCGAGGTGAGACGCGCTCCCTATTACCGCG 152  
13500 GCTGTCGAGATCATCTGACAGACGAGGTGAGAGGAGTCTTACCTTCCGCGC 13441  
153 CTTTCTGAGGCGCTTCCACCCCTGAGGCGCTGCGCGGCTTCCCTGAGAGAGTCT 212  
13440 CTTTCTGAGAAATGCGCCGATGTCGAGGCGCTGCGCGCGCGCCGACGAGAGAGTCT 13381  
213 TAGGCTGTCGAGGGGGGGGCTACTACGCGGGGGGGAACACTTCAACGCGCTGCGCG 272  
13380 GAGGCGCTGCGGGGCGGCTGCGCTATTCTCCGCGCGCCGACCTTCAAGCGCTGCGCG 13321  
273 AAGCGTGAAGAGC-----TTCCCGGAGCTTTCGCGAGCTTTCGCGGGCTTC 320  
13320 TCTGCTGGGGGGCGCGGCGTGGCGGCTTCCGATACGAGGCGGATTTGAGAGACTGCG 13261  
321 TGGCTCGGCGCTTACACCGCGGGCGGCTTCACTGCTTTCGCGGAGAGCGGTGCG 380  
13260 CGGCACTCGGCGCTTACACTTTCGCGCGGATCAAGGCGGATCGCTTTCGAGCGCGCGCG 13201  
381 GCGGCTGAGCGGGAAGCTCCGAGGCTCTCCCGCTCTTTCGCGCGGAGAAAGCCCA 440  
13200 CTTGCTGATGCGAATGTGAAAGCGCTCATATGAGGCTTTCATCATGAGAGCGCACT 13141  
441 GG-----AGAGAGCTTTCGCGCTTCGCGCAAGGCTTTCGCGAGGCGCTGCA 491  
13140 CAACGAGGCAAGACGAGATGAGAGCTTTCGAGGCGCATGCTGCGCGGCAAGG-- 13083  
492 CCGCGGAGGTGTGAGACGAGGCTTTCATGAGCTTCGCGGCGCAAGGCTGCTGCGGAAAG 551  
13082 -CCGCGGAGATTCGCGCGGCGCATGATGATCTTCGCGCGCAAGATGCAACCCGCGCG 13024  
552 GCGCGGCTGCGGCGCTTCGCGCGCTTTCGCGGCGCTTTCGCGCGG 592  
13023 GCGCGCTGATGCTGCTGCGCGCTGCGCGGAGAACTGCAAG 12983

RESULT 12  
AF121797 1106 bp DNA linear BCT 21-FEB-2001  
LOCUS Streptomyces antibioticus A/G-specific adenine glycosylase (muty)  
DEFINITION  
gene, complete cds.  
ACCESSION AF121797  
VERSION AF121797.1 GI:4455097  
KEYWORDS  
SOURCE Streptomyces antibioticus  
ORGANISM Streptomyces antibioticus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycinae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 1106)  
AUTHORS Jones, G.H., Paget, M.S., Chamberlin, L. and Butner, M.J.  
TITLE Sigma-E is required for the production of the antibiotic  
actinomycin in Streptomyces antibioticus  
JOURNAL Mol. Microbiol. 23 (1), 169-178 (1997)  
MEDLINE 97157951  
PUBMED 9004230  
REFERENCE 2 (bases 1 to 1106)  
AUTHORS Bralley, P. and Jones, G.H.  
TITLE Transcriptional analysis and regulation of the sigma-E gene of

Streptomyces antibioticus  
Biochim. Biophys. Acta 1517 (3), 410-415 (2001)  
MEDLINE 21240331  
PUBMED 11342219  
REFERENCE 3 (bases 1 to 1106)  
AUTHORS Bralley, P. and Jones, G.H.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUN-1999) Biology, Emory University, 1510 Clifton Rd,  
Atlanta, GA 30322, USA

## FEATURES

source

1..1106

/organism="Streptomyces antibioticus"

/mol\_type="genomic DNA"

/strain="IMRU 3720"

/db\_xref="taxon:1890"

/clone="pUSE1101"

183..1106

/gene="mutY"

183..1106

/gene="mutY"

/function="DNA mismatch repair"

/note="mutY"

/codon\_start=1

/transl\_table=11

/product="A/G-specific adenine glycosylase"

/protein\_id="AAD1076.1"

/db\_xref="GI:4455098"

/translation="MTAPTKPSSPADVDVSGADLHPVIDMPDHDARDCPWRPEAG  
AMAAVWSEFMQTPVSRVLYVYRWQLARVPRDLADAGAEVAMRGYRRAL  
RLHGAVALTERHGGDVGTQLALPGEYTPAAVAVSPAYQGRHVALDTNRRV  
ARATGVQVPPNATTAERRLARLALPEDEGTARWMAAAMELGALVCTANETCVRC  
PIAGOCAMRLAGKREHGGPPRGQTVGTVGRVSKLVALVREAHAVPQBALDQWQ  
EPVQPPRALDGLVDPGLVPLRGYRLPLT"

## ORIGIN

Query Match 13.3%; Score 129.4; DB 1; Length 1106;  
Best Local Similarity 51.8%; Pred. No. 3.4e-08;  
Matches 413; Conservative 0; Mismatches 351; Indels 33; Gaps 4;

20 CCTCTCTGCTGTGACCGGAAAGCGCCCGCTCTCCCTGGCGG---GGGGAGAAAG 76  
253 CGGTCATGACGTGTTTCACGACCAAGCGCGCTGATGCTCCCTGGCGGACCCCGAGGCGG 312  
77 ACCCTTACCGGCTCGGTCTCCGAGGTCTTTCGACAGACCGGAGTGAAGAGCGCC 136  
313 GCGGCTGGCGCTTATGCTCAGAGTTCATGCTCAGACAGACCCCGGTGAGCGGGTCC 372  
137 TCCCTATTTCACCGCGCTTTCGAGCGCTTTCACACCTGAAAGCCCTGCGCGGCTT 196  
373 TCCGCTTACGAGAGTGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGC 432  
197 CCTGGAAGAGGTCTTTCGAGGTTCGAGCGCGGCGGCGCTTACACCGGCGGCGGAGAAC 256  
433 CCGCGGAGAGCGCTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 492  
257 TCCACGCGCTGCGCGGAGCGTTCGAGAGCTTTCGCGGAGCTTTC-----G 301  
493 TCCAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 552  
302 CCGAGCTTTCGCGGCGGCTTTCGAGGTTCGAGCGCGGCGGCGGCGGCGGCGGCGGCGG 361  
553 CCGAGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 612  
362 CTTTGGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 421  
613 CTTAGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 672  
422 TCGC-----CCGGAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 469  
673 TACAGGCGGTGAGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 732  
470 GCTCTCTCCGAGGCGGTGACCGGCGGCGG---TGAACAGAGCGCTTATGAGCTTG 526

[illegible]

RESULT 13					
AE005710	AE005710	11283 bp	DNA	linear	BCT 12-JUN-2002
LOCUS	Ca10bacter crescentus CB15 section 36 of genome.	359 of the complete			
DEFINITION	AE005710				
ACCESSION	AE005710	AE005673			
VERSION	AE005710.1	GI:13421519			
KEYWORDS					
SOURCE	Ca10bacter crescentus CB15				
ORGANISM	Ca10bacter crescentus CB15				

REFERENCE	AUTHORS
1 (bases 1 to 11283)	Nierman, W.C., Feldblyum, T.V., Taub, M.T., Paulsen, I.T., Nelson, K.E., Eisen, J., Heidelberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R., Pocock, A.I., Nelson, W.C., Newton, A., Stephens, C., Phakee, N.D., Ely, B., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwim, M.L., Hatt, D.H., Kolonay, J.F., Smith, J., Craven, M.B., Khouri, H., Shetty, J., Berry, K., Uetzbach, T., Tran, K., Wolf, A., Vamathevan, J., Ermolova, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L. and Fraser, C.M.
TITLE	Complete genome sequence of <i>Caulobacter crescentus</i>
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
MEDLINE	21173698
PUBMED	11259647
2 (bases 1 to 11283)	
REFERENCE	Nierman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J.,

TITLE	Complete genome sequence of <i>Caulobacter crescentus</i>
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
MEDLINE	21173698
PUBMED	11259647
REFERENCE	2 (bases 1 to 11283)
AUTHORS	Nierman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J., Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R., Potocka, T., Newton, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B., Laub, M.T., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Berry, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouli, H., Shetty, J., Hartz, K., Utecherback, T., Tran, K., Wolf, A., Yamanehvan, J., Ermolova, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and Fraser, C.M.
TITLE	Direct Submission
JOURNAL	Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
FEATURES	Medical Center Dr, Rockville, MD 20850, USA
SOURCE	Location/Qualifiers
	1. .11283

gene	complement (254 . . 763)	/gene="CC0365"	complement (254 . . 763)	/gene="CC0365"	/note="identified by match to PPM protein family HMM PF00430"	/codon_start=1
CDS						

```

gene
    /transl_table=1
    /product="ATP synthase F0, B subunit"
    /protein_id="AAK22352.1"
    /db_xref="GI:13421520.1"
    /translation="MEHQSLTFSNPFEFWIALVIFEGLLIVLKVLPGLPGALDGY
AAKIKAEIDEMEQQLREEQALADLVKQORBDPEAQAAAMEAQAADAKRIAEKEKEL
EEOQLRRREMERKATQAEQAQADVAADVAALAAQAEFTVLAIRLAGAKDITLVDAAL
IGMGAKXIO"
    complement(1774. .1307)
    /gene="CC0366"
    CDS
        complement(1774. .1307)
        /gene="CC0366"
        /note="identified by match to Pfam protein family HMM
PF00430"
        /codon_start=1
        /transl_table=1
        /product="ATP synthase F0, B' subunit"
        /protein_id="AAK22353.1"
        /db_xref="GI:13421521.1"
        /translation="MATEHHGTETETEPAPKHESGGLPOLQPEHWGGQIVMLLIIFAQVTV
YAVSKALLPVSQAIDBERGKKTIGDILADARHMDQEAQAARAAAVAEARAEQVTE
AADAAKASASAAERQAEAEAVLAEKLLAAEASIQIARDDMSQVRVAETGATVCE
KLGTGASAAALKCALA"
        complement(1434. .1658)
        /gene="CC0367"
        CDS
            complement(1434. .1658)
            /gene="CC0367"
            /note="identified by match to protein family HMM"
            /codon_start=1
            /transl_table=1
            /product="ATP synthase F0, C subunit"
            /protein_id="AAK22354.1"
            /db_xref="GI:13421522.1"
            /translation="MDAAAARYIGAGLAMIQMGIGAGVGLGMEGNYOGALRNPTAAA
QERPMFLPGMALTEALGIFALVIAFILFS"
            complement(1708. .2469)
            /gene="CC0368"
            CDS
                complement(1708. .2469)
                /gene="CC0368"
                /note="identified by match to protein family HMM"
                /codon_start=1
                /transl_table=1
                /product="ATP synthase F0, A subunit"
                /protein_id="AAK22355.1"
                /db_xref="GI:13421523.1"
                /translation="MADPMHQFOIQIVLEPTVTPGLAIDLSITNSVAAAMSALLI
IIGFAALSAKKAIVGRLOSQSGEMLYGIVDLAESIIIGHDKKFLPEIPTLFAVLPE
NVQGVLEFTRTASOLAVTLTGLWVILTVVAVGAKNGLKPFKLFAASGVMPYPTFLI
VPIELSLAIPITLIGRLFGNMLGSHVLIKIPGFVALGGLGVLMAGAAALVTSV
VALTLERFVAFLQAFVFAVLAICYINDVHLDHS"
                complement(2475. .2855)
                /gene="CC0369"
                CDS
                    complement(2475. .2855)
                    /gene="CC0369"
                    /note="identified by Glimmer2; putative"
                    /codon_start=1
                    /transl_table=1
                    /product="hypothetical protein"
                    /protein_id="AAK22356.1"
                    /db_xref="GI:13421524.1"
                    /translation="MPKADENDAKIESDARLAAFAFKKAAEKSKATSEKATODQDY
RLIALDLGGVGVGGFWGLDHYVATSPWMTIGLLIGLGATISYIRKAMKLSAQMS
APSPGNAGDEDTGSGTAPKQKRD"
                    complement(2854. .3222)
                    /gene="CC0370"
                    CDS
                        complement(2854. .3222)
                        /gene="CC0370"
                        /note="identified by Glimmer2; putative"
                        /codon_start=1
                        /transl_table=1
                        /product="hypothetical protein"
                        /protein_id="AAK22357.1"
                        /db_xref="GI:13421525.1"

```

gene  
CDS  
3277. .3666  
/gene="CC0371"  
3277. .3666  
/gene="CC0371"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK22358.1"  
/db\_xref="GI:13421526"  
/translation="MSHLKTYGPAIDRISAGQEPKAMIAKRPKPTVQDDFDKREEQ  
RAALEVAVARAARAEKARLEALRAOEOMAVKAEKREKATIEAEOBMRKEEKAK  
EBRELRLGEPNASKASRAHQWASLIG"  
3710. .4516  
/gene="CC0372"  
3710. .4516  
/gene="CC0372"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK22359.1"  
/db\_xref="GI:13421527"  
/translation="MAERDEPVMVSLDIASLIPVVALAGALDAINGGGILITPLAL  
LGLASPVQALQGNKIQCHISALSTSPAFARGLIDMKTLPAVGAALGLGALCA  
SLSPFLAAVPLMLIAIALYFGLSRAIKEDVTPRALIPKAPVPLIGYDDI  
GPGAGFVWVAIVTLGALVATHTLNAASNLGSLVFTLGKVVWVQLWMAA  
GATIGAVQSVLAVMRGPKLIRPLLVISCAMVAILDLVDPANIPLMALIG"  
complement (4631. .8074)  
/gene="CC0373"  
complement (4631. .8074)  
/gene="CC0373"  
/note="identified by match to protein family HMM"  
/codon\_start=1  
/transl\_table=11  
/product="emc protein"  
/protein\_id="AAK22360.1"  
/db\_xref="GI:13421528"  
/translation="MQFORLILSGFSKPVETERTIRPGITGIVGNGCGKSNLLEAL  
RMYGNSAKKARAGMDVLPASGAPAPNNAVDVLTIDNADRAPOFNDPILB  
VVRILDRGGSTYRINRGVAVRVQLIPDASGANSPALVQOQISELIGAKPNR  
RLTBEEAGVGLTRHREAEILRLAAETINSLREDAVARELETLARLREAAQEKY  
KLSLEIAVQGVAVYARMTBARDHLETTSEATPAARLVEETARASAAVAITAE  
AAMPIREBATTAGAILGQALIQDRAEREEAAMAEFRLSNDLSRIDADAREQA  
KODAAALARIAPLEBEERVALVAAAPRGPIAAVAAAEBAAMAAVEQALAEQA  
AEAGGRRARLSEAEKARNRTTRALEARQAEPAAGPEVPDAADAROFANEEA  
LAAARALBEASTARKVAEOEQAORQLASVEQDQLRTERGLAQUTAPRSKSH  
APALDPSVDKGYGALAAALGDDLDALDPKAPSYWGAEAPAPVMPGASPLAPLV  
KAPPLAARLSHVAVVTTPANGDRLOKELKGMRLVSGDGLNMGDFVAPRAAPAA  
VLTREORTLAEVEAIDVMAAPRAATTITAKAADRRAAEDGLRKRKGPDAEREL  
TOAROVKLEFEEOALRAARQASLDITIGRPKAEKADALGEAEAAHAAQSGDL  
QGLEAEOALQAEAAAGARTLDVETREARQRRLSELRDDADMSKTAEEAAK  
BRESIEGRVYAAAALEAPARAPALOEKVALUDEPVALAEARAAVASALETAETTR  
LMAADRAAPAEQAGEAREKKAVALHLDGARPAEVAVALREQMRSEBELGATTA  
GGAIVPDDAGVEARHLFALERERDAPGPNVLAEEBAEQVAYARLETMSSEADLSGA  
YTKLAGAIEELNAGBERLLAAPDVINANQTFQALFGGQAEELTIESDDPLEAG  
ELFAPCPKPRKMSMSINGEOLTAGALIFGVFLANPAPICVLDVDAPLDADAVR  
YCNMLDERRRRTQTRFATITNPVTMSRMDRLFGVIMABRGVSQALVSVDLSAEKLV  
A"  
complement (8096. .8704)  
/gene="CC0374"  
complement (8096. .8704)  
/gene="CC0374"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK22361.1"  
/db\_xref="GI:13421529"  
/translation="IMTLDRILILASGLATAGSGVARSRLPPAPGDVNLGAATAPVO

Query Match	13.1%	Score 127.8	DB 1	Length 11283
Best Local Similarity	53.6%	Prod. No. 2.7e-08		
Matches 351	Conservative 0	Mismatch 277	Indels 27	Gaps 3
gene	<p> FVVAASPSGCHCAHMTTELDAIKRTPLDGRKLVLRRELTLPNEFAANGFLIARRV  PKRATREVLITFQKRETFESGRLMBSGLQALADQYGLTEAQTFTAMNDITALDGVNTR  PRAATGEDQVEVETPLFFVNGAPRYEGEADLALSKIFALAAG" </p>			
CDS	<p> complement (8704) .,9318  /gene"CC0375"  complement (8704) .,9318 </p>			
Qy	55	CTCCCTGCGCGGGGAGAGAAGACCCCTTACCGGCTCTCGGCTCCGAGGCTCTTTCGAC	114	
Db	10234	CCCCCGGGGGGCGTCCGACGATCCCTACCGGCTCTGCTGTCGAGGATGCTGACG	10293	
Qy	115	CAGACCCGGGTGAGACAGAGCCCTCCCTATTACCGGCTTTTGAGCGCTTTCCACC	174	
Db	10294	CAGACCAAGGTGACCCACAGCCACGCCCTTATTTCTTAGCTTACACCGGCTGGCCGAG	10353	
Qy	175	CTGAAGGCGCCCTGGCGCGGCTTCCCTGMAAGAGTCTCTTAGGTCTTGACAGGGGGGGC	234	
Db	10354	GCTTAGACCTTCGGGCTGTGAGACGGGACCTAGTGGCCGCTGGGGGGGCTTTGGC	10411	
Qy	235	TACTACCGGCGGGGGAACACCTTCACCGCTGGCCCGAAGCGTGAAGAGCTTTCCCGC	294	
Db	10414	TACTACCGCCCGCCCGCACTCTTGAGCTGTGCGCGGGTGGCAACATCAGCG	10472	
Qy	295	AGCTTC-----GCCGACCTTCGGGGCTTCTGTGCTTCGGGCTTACCC	339	
Db	10474	GCGCTCTTTCCGGGACCCGAAGAGGCTCGCGCTTCGCGGGCGTGGGGGCTTACCC	10533	
Qy	340	GCGGCGGCGGTGAGCTTCATCGCTTTCGGGAGCGGAGTGGCGGCGGTGACGGGAACGTC	399	
Db	10534	GCGCGCGCGTGGCGGGATGCGCTTGACCGGCGCGCAATGTCTGACCGCAATGTC	10593	
Qy	400	CGAGAGGCTCTCTCCGCTCTTTCGCCGGGAAAGCC-----CAAGAGAAGAG	450	
Db	10594	GAGGGGAGTGTGCGCGGTTCGTGTGAGAGCGCCGATGCCGACAGCAAGCCCGAG	10653	
Qy	451	CTTTCCGCTCGCGCCAGGCGCTCTCCCGAGAGGGGTGGACCCGGGGGTGAGAACAG	510	
Db	10654	TTGAAGGCTCTGGCGCGGCGACCTGTGACCGA---CGATCGCCCGCGACCTGGGCCAG	10710	
Qy	511	GCCCTCATGAGCTCGGGGCGACGCTGTGCTTCGCGAAGCGGCCCGTTCGGGGCTTCG	570	
Db	10711	GCGCTGATGATTTGGGCGCGACCGCTCGAAGCCCAAGGCTGCGGTGCGACCGCTGC	10770	
Qy	571	CCCTAGGGGCTTTCGCGGGGGAGAGAGGCCCGCGGCGCTACCCGCGCCAGAAAG	630	
Db	10771	CCGGTCTCGCTCTGTGTGCGGGCTTTATGTGGGGGGGGCCCGGAGACTATTCGAGAAAG	10830	
Qy	631	CGCCGGCGGAAGAGAGCGCCTCTGCTGCTCTTCGCGGCGGAGAGGGG	685	
Db	10831	ACCAAGAGGCGGACCGGCGCGGCGAGGAGCGGGTGGCCTATGTCTACGCGCG	10885	
RESULT 14				
EX572607/c				
LOCUS	EX572607	345012 bp	DNA	linear
DEFINITION	Rhodopseudomonas palustris CGA009 complete genome; segment 15/16.			
ACCESSION	BX572607 BX571963			
VERSION	BX572607.1 GI:39651254			
KEYWORDS	complete genomes.			
SOURCE	Rhodopseudomonas palustris CGA009			
ORGANISM	Rhodopseudomonas palustris CGA009			
REFERENCE	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Rhodopseudomonas.			
AUTHORS	1 (bases 1 to 345012) Larimer, F.W., Chain, P., Hauser, L., Lamedin, J., Melletti, S., Do, L., Gibson, M.L., Pelletier, D.A., Beatty, T.J., Lang, A.S., Tabita, F.R., Land, N.L., Hanson, T.E., Torres y Torres, J., Peters, C., Harrison, F.H., Gibson, J., and Harwood, C.S.			
TITLE	Complete genome sequence of the metabolically versatile			

JOURNAL photoresynthetic bacterium Rhodospseudomonas palustris  
REFERENCE 2 (bases 1 to 345012)  
AUTHORS Larimer, F.W. and Harwood, C.S.  
CONSTRM Rhodospseudomonas genome consortium  
TITLE Direct Submission  
JOURNAL Submitted (24-JUL-2003) Submitted on behalf of the Rhodospseudomonas  
genome consortium, the DOE Joint Genome Institute, Production  
Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,  
USA, and the Genome Analysis Group, Oak Ridge National Laboratory,  
1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;  
larimerf@ornl.gov

FEATURES  
source

1. 345012  
/organism="Rhodospseudomonas palustris CGA09"  
/mol\_type="genomic DNA"  
/strain="CGA09"  
/db\_xref="taxon:258594"  
280..1221  
/locus\_tag="RPA4336"  
280..1221  
/locus\_tag="RPA4336"  
/function="InterPro IPR001395 COGS COG0667"  
/note="observed by proteomics  
Citation: Proteomics from Verberkmoes et al. (2003)  
unpublished"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="putative oxidoreductase"  
/protein\_id="CAE29777.1"  
/db\_xref="GI:39651255"  
/translation="MDIRNLGSGSLRVSVGLGNNFGQRTDPTSKVTHKALDLY  
TLPTADIVAGMGSETEVLGDRKDIYATKPKMATDGTGQASRRYVEAV  
EASLKRKTMDIDYQOHDPLPTIDETLALDDLIRAGKVRIGNSPPAMRIATA  
EYVARALGTRHYVSCDESYIVVRDIEKILLPAQEXKGLLPFPPLASGLITGKYR  
GTEAPATRRAKMADIRDRYTTDANLIDVTKKPARARGSLLELAFMSLACRPQS  
SVAGATTPEQIEONRAVSWKLTAEEMAEVDALTKG"  
1671..2414  
/locus\_tag="RPA4337"  
1671..2414  
/locus\_tag="RPA4337"  
/function="InterPro IPR005834:IPR006402 COGS COG0637"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="putative phosphoglycolate phosphatase"  
/protein\_id="CAE29778.1"  
/db\_xref="GI:39651256"  
/translation="MSARGCEGEPQVVTGAHDMLIAAVLLDMGTLDVTERLYLESLT  
EVLNAPGLPDAIFETCESWVGLPGEQCALVARYGETLPLEINRAFEKKDAPAG  
LPVAPGTCELDLDADGAPVAVVTSSSRKADNHLTAGIRARFATILTEDDVVHGK  
PAPDLVLAANRLGVPAHCVAVEDSSGVSAPTAAGITLMVDDLQPDATRGKCA  
AVLPDLNAYRSLLEIRGRFAAPDAALAPR"  
complement (2341..3021)  
/locus\_tag="RPA4338"  
complement (2341..3021)  
/locus\_tag="RPA4338"  
/function="InterPro IPR001647 COGS COG1309"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="transcriptional regulator, TetR family"  
/protein\_id="CAE29779.1"  
/db\_xref="GI:39651257"  
/translation="MAVRVSVAPKAPPPSPTRDRIVAANAALFYSEGRANGVNAE  
KAGVTKTLTYHFRSKODLVNAVYEGRQPVLAIFROMFESTPGLPKVIGTRINLA  
RAAHPKMGCGFLRTSAELANLPQHPIKIGAAKKKFERMLRETFEAGITADEL  
ARQWLLLDGSAFVAVLLHRDPSYMETGAAAKLVAAALPRSFNAARPPAPRTGGS  
SRANSGRP"  
3103..3894  
/locus\_tag="RPA4339"  
3103..3894

gene  
CDS  
/locus\_tag="RPA4339"  
/EC\_number="4.2.1.17"  
/function="InterPro IPR001753 COGS COG1024"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="possible croconase, 3-hydroxybutyryl-CoA  
dehydratase"  
/protein\_id="CAE29780.1"  
/db\_xref="GI:39651258"  
/translation="MTELEIPELRDGVALLTNRPEKNAISVALIDRLMQRDAIEV  
DNAVRVILTAGSERAIFSAAGADIAEPSASVARGADAVRQFVRGQMTARLSFKP  
VIAAVNGAYGCGCEITEAVHLAVASDRAPFAKDEILAMPPTFGTORLPRLGRKR  
ALFLITGEPSPHALAIGLVNRVYVHEELIGADALAIQIIRHPGVAVAATATT  
RGMLSTIDEGELVERAOFMLAGREBELSHLAPRRNRPRDAQTG"  
complement (3882..4685)  
/gene="119E"  
/locus\_tag="RPA4340"  
complement (3882..4685)  
/gene="119E"  
/locus\_tag="RPA4340"  
complement (3882..4685)  
/locus\_tag="RPA4340"  
/function="InterPro IPR004045 COGS COG0625"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="putative ligin beta-ether hydrolase"  
/protein\_id="CAE29781.1"  
/db\_xref="GI:39651259"  
/translation="MDRRIRKSGDVRGCRGPAPDMCECPQSRDLTRDMTKLFL  
VGDAGRFPSPFWRTMALAHKGLTAEITPMKFTDPAALAPYSEKVPVLLDDKVV  
ADSWALNYVEDAYPDRSLFEGGCGGAMRMNMDLAIIGIIFMIVADLRIQD  
PDAAYVROSREARLKTLEEAASRPRAVQGRKSLDPLRLTKQAFIIGAANPVA  
DYIVFGSGFQARVVSDFSLQSDDPYVAMERLLDARGLARATPGIPV"  
complement (4874..6859)  
/locus\_tag="RPA4341"  
complement (4874..6859)  
/locus\_tag="RPA4341"  
/function="InterPro IPR001279 COGS COG2015"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="Beta-Lactamase-like"  
/protein\_id="CAE29782.1"  
/db\_xref="GI:39651260"  
/translation="MARIIPSMILATTAALSLAAPLAQPNDAEPATRSANVYNS  
LPLADLADGFDAQRGLIASLPDGVPPGAPAPAMDLKPYDFLGDPSPATVNSLR  
QAQNLNLSGGLFOVAERYQVRGLDIANVTIVEGDTGLINDTLTVGTAAKALDLYQ  
HRPKKPVALLIYHSHIDHFGARGLIDEADAAGKVAAPGLFLEHAVAENVLAGN  
AMRRAPQFORGTOLPVGERGOVDAGLKAGVSLIAPNDLIKQYETRSDIGVEI  
EPIHVPSEAPSENIISYPOFKVLNMAEDTHTLNLTYLRGAAIRGRLMSKYIGEA  
IERGDDTVDVIAQHMNPVGRDRVGYLKKQSDVKFTIDQSRLLNHGILPTETIAB  
RLTLPBLTSEFARAGYGVSHNAKAVQFYLGWDANPADLNPILPRAQAKETIDY  
MGGAAYLIAARDPKAGQRYAVATVASKLVFIDPATTEPRLAGADLEQGYOAEBS  
TWRNAVYLGAQELRNLKIDTSVTSNDELKVFIDAPFEPLARLNAAKEGGHIV  
NMTFTDKERTYTNLNLNSALSTHISGKSDNADSVTLNFTFDALISLKQGFAGVLS  
GDLWVSGNPLKSLKELFEDFSPNFEVIEPVAKAKE"  
complement (6782..6859)  
/locus\_tag="RPA4341"  
/note="Signal predicted by SignalP 2.0 HMM (Signal  
peptideprobability 1.000) with cleavage site probability  
0.916 atresidue 26"  
7091..8143  
/locus\_tag="RPA4342"  
7091..8143  
/locus\_tag="RPA4342"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="CAE29783.1"  
/db\_xref="GI:39651261"  
/translation="MRYSGEDLQAAAGVITPAQEAALPAPIDARRPAAASDAPRD  
IVHLWYAGALIVISAMGLFSTMAFAGLGAALATATLTIYAVAFWMTGHLVHYKNLR



/codon\_start=1  
/transl\_table=1  
/product="Vng1510c"  
/protein\_id="AAG19804.1"  
/db\_xref="GI:10581002"  
/translation="MSSNADGGAFOAACELVSRILAGVSKDDIEDAKLDGCEYAA  
SKVPHSDLDLVAAPDGRDELVDLQKRPVATAGVPAVIMTSPHCCKYCPG  
GPAEPTSSOSYTGHEPAAAGVONDDYDPGVQVTLRLHQLRGHPVQKVELIMGGT  
MTASHDYQEFVYKALQALNDPDPDEPAPNQTSPADPAEDFRLIEVIANDEH  
GALKNIGTFETKPDWCDEQVDRMLDGAATKVVGQVTFERVREMERGHVQASV  
DANRLRDLAAKVGFMHPGQGMSEKENVLEDPRRLPDERMKPDYLKIPTLVVRDT  
ITYDMRNGDEPFLPHEADLVAEIKMIPRYRLQPRQDIPADFIADWMSNLR  
OLARORAEHGYTCDCIRCREAGNDDPEVSLDVETAAAGGTEQFISIEDQDL  
LVGFARLRFPRPAPARELTDAALRELVHVSSEVSGDDGDDQHOGYGRQLEMEA  
EHLAADSGFEKTSYISGIGARQYKQQLGHQDPYYSKRL"  
complement (1767. .3821)  
/gene="VNG1511C"  
complement (1767. .3821)  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/product="Vng1511c"  
/protein\_id="AAG19805.1"  
/db\_xref="GI:10581003"  
/translation="MQMATCIYCCSSTDGHCATHEEDVVPDFRDSPPDQITPGRFPR  
GTVDGFAPFVGVVGGTVTGLHRSVPKGVDSLAMVGDVVFQVVDVHNDNIDL  
GWSIRBSMRDFRGLVHDPANDSALPADADSDTDDGTTDSDADATADGKR  
VSDADATTERAADASFESGADDDDTHEITVRVDALDDHVDADVTEGEITAAQTSQ  
PYFELTDETCADVAAAFVAGVAPRIEFGAVVVLGTVERRGDLQVTEAVTEL  
TGDRADVRNMDALDDRAAPATALLAPDDADVAAHDDIYDAATVRAVAVDAP  
VTVNPTTVSGYVAGTAIERALLPLIRDEHAREDAEHYVDVRLDAPFYITDADGD  
VTSMLAAHRHDERKPLFVFGSTSTESTDALDLIDYADTVVVDGVDADAAGMD  
VLVSPTAGSEQPVTGALQGLAAVINDVREDIAHLPAVAVWTDLPAAVADLAADS  
YDSARALRDVALLEAFYOSGEDKRELDLADGDDADLADHIAIQFRRLATELS  
TAEPHLSTRHHDGVAPELIDVGYATHOVDPFPVLLLDALHRAADSTDVLAGEDNL  
RVRSDDAVDVAAVDAIDDDLTITTAGTARAKARGRLEFLAGERDAVLDAVVDAAQ  
LA"  
complement (3867. .4103)  
/gene="VNG1513H"  
complement (3867. .4103)  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/product="Vng1513h"  
/protein\_id="AAG19806.1"  
/db\_xref="GI:10581004"  
/translation="MVELKDSIKVNGPTNKHADKLFDVUDATAREGDKRVLVFE  
DDSGNEVQVALFPEQVAAALADDIDALTRDSVFE"  
complement (4154. .4762)  
/gene="VNG1514H"  
complement (4154. .4762)  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/product="Vng1514h"  
/protein\_id="AAG19807.1"  
/db\_xref="GI:10581005"  
/translation="MTTWPEBEGGREREPGLGLASTFOVTLNPVSPFDEAVSPDDAP  
GLVAPMVVGVAAATELLATPESDALAPPTADMLAVALVLAIVLLVTPAAPAAHALSAIQ  
TLCLIVATISBSRGVSETVQLIYAAAPCVVTIGVAVLTLGLGVAMCCLLVGTSTV  
VHSLIPRAVIAAAPVIAFAGAGFCAIPAGEQYIHAAGLAF"  
complement (4766. .5353)  
/gene="tdk"  
/note="synonym: VNG1515G"  
complement (4766. .5353)  
/gene="tdk"  
/note="tdk"  
/codon\_start=1  
/transl\_table=1  
/product="thymidine kinase"  
/product="thymidine kinase"

/protein\_id="AAG19808.1"  
/db\_xref="GI:10581006"  
/translation="MHAITNSGWJEVLTGMSFKTEELLRLRRAETAGODVAATP  
AVDRYGEATLGSNAGSMATVVEPTAREGASIPTLINGQVVAIDANFPPELVVD  
VCOBLADGRRVVLSTGDTQTFRGEFPFPPQMLAIEYVDKRAICMQCGSPATRNOR  
LIEGEPAHYDDPTVMVAEETFEARCRNHVRE"  
5460. .5900  
/gene="VNG1518H"  
5460. .5900  
/note="VNG1518H"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/product="Vng1518h"  
/protein\_id="AAG19809.1"  
/db\_xref="GI:10581007"  
/translation="MYDDILLPTDGSADAANAVIDLASKYNATLHVLYADTTEY  
STVTFEADVTPLEBEGERVVDVAABEADGVAMGVHLGGVFETIOAYVDEGID  
TVWGTGRRGLGALLGSVTERVVRTSDVPLTVRETFDDGAA"  
complement (5917. .6126)  
/gene="VNG1519H"  
complement (5917. .6126)  
/note="VNG1519H"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/product="Vng1519h"  
/protein\_id="AAG19810.1"  
/db\_xref="GI:10581008"  
/translation="MEHNVGMDALRLGLGAVLGSAAATLACVGGSPALTPALGV  
GALLILASGMNTGGVRYLVGVDTT"  
complement (6203. .7141)  
/gene="mut1"  
/note="synonym: VNG1520G"  
complement (6203. .7141)  
/note="mut1"  
/note="mut1"  
/codon\_start=1  
/transl\_table=1  
/product="A/G specific adenine glycosylase, repair  
protein"  
/protein\_id="AAG19811.1"  
/db\_xref="GI:10581009"  
/translation="MTTGDSGRAGACGPADTTALQTLADWVYTDSHRSFPWRETTDP  
YEILVSEVMSQQTQLSRVIDAWRAFLLRWPTTALAADRDVGVFWSHSLGYNRA  
THLHEAQOVEDTDGAIPTPTAPLSEISMGVPTANAASFAFNAGNAVVDVTKRV  
LYRAFEGIRDDDDPDYRPLANELLPDGSRVMNNAVMEIGAVACQCFRCDEAECPFR  
EMCHAYOTGDPTADVPVTPDSFEGSRQFRGRIVRLIGHEDEMELDALGHRVVRDYTP  
DGEYGRMGLSLDLDLDDGLVRYEYGERIVVBLE"  
7627. .9558  
/gene="Htr8"  
/note="synonym: VNG1523G"  
7627. .9558  
/gene="Htr8"  
/note="Htr8"  
/note="Htr8"  
/codon\_start=1  
/transl\_table=1  
/product="Htr8 transducer"  
/protein\_id="AAG19812.1"  
/db\_xref="GI:10581010"  
/translation="MGSNARGGSAETTGGLAAVKTYYLETPTCGESIPAAWRRHR  
NVRVFLAHLPLILATLALYGTESAVTGATIPPTGILIAELGIVGLVGLASIPSV  
SRRRFTLASITVAASVVAVQFSGFIKAHFHFGMAVAVAYEDMLPFLGLGVV  
FTHGFGMIAERVYNTTAAINNPMWGGIHGAFVLILAGALAMWYSTERASRSOK  
RLBEARQAOQVEDLEARQAEIEAEKAEARKADAEAREAAAOOREVALNRIE  
ATANTYGAARAAVADGLSVRLDDVDVNDMAALYAAFPNEMLDETETTIREIOVASD  
VAAASEPADAGVETIEDASQVSETVVEIAGADEQEKLETGSGMTDLSAIEEVA  
ASADSVAERSHETNAVAGDSQTRBQALIDSRTVQSVESTVQVVEALDDQIAETSI  
VDLISDAVEQTNNMLANANI EAARADSGGFVAVADVEVOLAETRTGVOEISTTTDQ  
AIDAKQATVTEARTADESVQDAISADVAVDAFGVAAENAEETDGTGVOEISTTTDQ  
AASTEBAVSMIAEVSIDISTATAADAQAQASTAEOQTAAATISNTAALRQADRLQG  
LVSTFVDHDSASTAARSE"  
complement (9562. .10938)



CDS  
/gene="VNG1524C"  
complement(9562. .10938)  
/gene="VNG1524C"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/product="Vng1524C"

Query Match 12.7%; Score 123.8; DB 1; Length 11034;  
Best Local Similarity 53.9%; Pred. No. 9.4e-08;  
Matches 313; Conservative 0; Mismatches 247; Indels 21; Gaps 2;

```
QY 7 GCTTGGCGGAAAGCCCTCTCCCTGCTGTCACCGGAAAGCCCGCCCTCCCTGCGCG 66
Db 7084 GCCCTCCAGACCGCACTCTGTCAGTACACCGACTCCACCGGTCTCCCGTGGCG 7025
QY 67 GGGGAGAGAGACCCCTTACCGCGTCTCGAGTCTCCGAGTCTTCTGAGCAGACCGGAG 126
Db 7024 GAGACCAACGACCCCTTACGAGATCTGATCTCGAGGTATGAGCCAGACCAAGCTC 6965
QY 127 GAGCAGGCGCTCCCTATTACCGCGCTTCTGAGCGCTTCCACCTGAAGGCTTG 186
Db 6964 TCCCGGGTATCGACGGCGGTCTCTGATCGGTGGCCGACACCGCGGTTG 6905
QY 187 GCGCGGCTTCCCTGGAAGGTCTTAAAGTCTG-----CAGGGGGGGCTTACTAC 240
Db 6904 GCGGCGCGCAGCAGCAGCAGTGTGGGTCTGTCCGACACAGCTCGGTAAC 6845
QY 241 CGGCGGCGGAAACACTCCACCGCTGCGCCGAGCGTGA-----GGAG 285
Db 6844 AACCGCGCACTCACTCAAGAGCGGCCCAACAGTGAACGACGACGAGCGCG 6785
QY 286 CTTCCTCCGAGCTTGGCCGAGCTTGGGGGCTTCTGTCTGGGCTTACACCGGCG 345
Db 6784 ATCCGCGGAGCGCGGAGCTCTGAGCTCATGGCGGTGGCCCGTACACCGCAAC 6725
QY 346 GGGGTGGCTTCATCGCTTCCGGGAGCGGTTGCGCGGTGAGCGGAACGTCCGAGG 405
Db 6724 GCGGTGGCGTCTTCTGCTTCAAGCGCGCAACGCGTGTGACACGACGTCAAGCC 6665
QY 406 GTCTCTCCGCTCTTCCGCGGAAAGCCCAAGAGAGAGCTTTTCGCTCGCC 465
Db 6664 GTGCTTACCGCGCTTCAAGGGATCCGGAAGACGACCGGATTAACGGCGCGTG 6605
QY 466 CAGGCGCTCTCCCGAGGCGGTGAACCGGGGGTGTGAACAGGCCTCATGAGCTC 525
Db 6604 GCCAAGAGTGTGCGCGACAGTCCCGGCTGTGAACACCGCGTATGAATCTG 6545
QY 526 GGGGCCACGCTGCTGCGGAAACGGCCCGCTTGGGGGC 566
Db 6544 GGGGCGGTGCTGCGCAGACCCCGCTGTGACGAGGC 6504
```

Search completed: March 4, 2004, 05:06:40  
Job time : 2827 secs



us-09-938-901a-1.rni

GenCore version 5.1.6  
(c) 1993 - 2004 Compuq

```
; Search time 94 Seconds
(without alignments)
5756.140 Million cell updates/sec
```

tagtccccctcccggaagca 975

**i.dues**

Parameters: 1365418

3

```
5A_COMB.seq: *
5B_COMB.seq: *
6A_COMB.seq: *
6B_COMB.seq: *
PCTUS_COMB.seq: *
backfile1.seq: *
```

redicted by chance to have a score of the result being printed, total score distribution.

# IES

	Description
-991A-6832	Sequence 6832, Ap
-991A-6886	Sequence 6886, Ap
-991A-6739	Sequence 6739, Ap
-039A-4788	Sequence 4788, Ap
-991A-6783	Sequence 6783, Ap
-574-1	Sequence 1, Appl1
-656-38	Sequence 38, Appl
-855-38	Sequence 38, Appl
3-840A-2	Sequence 2, Appl
3-840A-1	Sequence 1, Appl
-557-12	Sequence 12, Appl
-532A-3654	Sequence 3654, Ap
-000C-1710	Sequence 1710, Ap
-028-23	Sequence 23, Appl
-614-23	Sequence 23, Appl
-040-23	Sequence 23, Appl
-028-1	Sequence 1, Appl1
-614-1	Sequence 1, Appl1
-040-1	Sequence 1, Appl1
-661A-1143	Sequence 1143, Ap
-8-552A-1	Sequence 1, Appl
-734C-3	Sequence 3, Appl1
-009A-3	Sequence 3, Appl1
-352-856	Sequence 856, Appl
-002A-4	Sequence 4, Appl1
-010-4	Sequence 4, Appl1
-429-4	Sequence 4, Appl1

Sequence 5, April
Sequence 5, April
Sequence 5, April
Sequence 7, April
Sequence 22, April
Sequence 76, April
Sequence 1, April
Sequence 1,725, April
Sequence 2086, April
Sequence 1804, April
Sequence 2158, April
Sequence 1, April
Sequence 1, April
Sequence 1, April
Sequence 1, April
Sequence 11, April
Sequence 48, April
Sequence 7, April

## ALIGNMENTS

```

RESULT 1
US-09-252-991A-6832
: Sequence 6832, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196, 136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 6832
:
: LENGTH: 963
:
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-6832

```

Query Match	12.9%	Score 125.4;	DB 4;	Length 963;
Best Local Similarity	54.7%;	Pred. No. 2.1e-16;		
Matches 311; Conservative	0;	Mismatches 231;	Indels 27;	Gaps 2;

Oy	55	CTCCCTTGCGCGGGGAGAAAGAACCTTACCGCTCTGTCTCGAGGTCCTTCTGAG	114
Db	53	CTGCCCTTGCGAGCAGGGGCATCACCCCTCACCGGGGTGTGGGTCTCGAANAATCATGCTGAG	112
Oy	115	CAGACCCGGGTGAGCAGGCGCTCCCTATTACCGCGCTTTCTGGAAGCTTTCCAC	174
Db	113	CAGACCCAGGTGACACACCGTCTCGTTACTTTCAGCCGTTTCATGTGCGACGCTTGCCAC	172
Oy	175	CTGAAGCCCTTGCGCGCGGCTTCTCTGGAAGAAGTCTTAAAGGTCTGCGACGGGCGGGC	234
Db	173	GTCGAGGACATGCGCGCGCGCGCGAGGACGAAAGTCTTGCACTGTGTGACCGGGCTCGGC	232
Oy	235	TACTACCGGGCGGGCGGAACACTCTCACCGCTGCGCCCGAAGCT-----G	279
Db	233	TACTACACCGCTGCGCGCACCTTGACAAAGACGGCGAGATCTGTGTGAGACGGCGATCGG	292
Oy	280	GAGAGCTTCCCCCGAGCTTGCAGAGTTTGGGGGCTTCTGTGCTTCTGGGCTTTACAC	339
Db	293	GGGAGATTCCCCCGAGAGTGAAGAACTGCCCGAATGCCCGGATCGGCGCTCCAC	352
Oy	340	GCGGCGGCGGTGGCTCATATGCTTCTCGGGAGCGGGGTGGCGGGTGAACGGAAAGCTC	399
Db	353	GCTGAGGCATATGCGCAGCTGTGTGATGAGGGCTTGGCGGACCGAGATCTTGACGGCAACGTTC	412

OY	400	CGAGAGGTCCTCTCCCGGCTCTTCCGCGGGAAAGCCCGAAGAGAGAGACTTTTGCC	459
Db	413	AAAGGCGCTACTGGCGCGCTACCTGGGCGCAGAGCGGCTATCCCGGCGAACCAGAGTGCC	472
OY	460	CTCGCCAGAGGCTCTCCGCCGAGGCGCTGAGACCCGGGAGTGTG-----AAC	507
Db	473	AGGGCGGTGTGGGAAAGCCGCGAAGCTTCACTCCCGCAGCAGCGGGTCACCACTACAC	532
OY	508	CAGGCGCTCATGAGACTCGGAGGCTACGAGTCTGCTTGCAGAAAGGCCCCCTTGGGGGCC	567
Db	533	CAGGCGATGTATGAGACCTCGGCGGCACCTCTTCAACGCGAGCAACCCAGTGTCTCTT	592
OY	568	TGCCCCCTAGAGGCGCTTCTGCGGGGAA	596
Db	593	TGCCCGCTGATCTCCGGCTGCAGCGGCA	621

```

RESULT 2
US-09-252-991A-6886/c
; Sequence 6886: Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6886
; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6886

```

Query Match	Similarity	12.9%	Score 125.4	DB 4	length 1401
Best Local	Similarity	54.7%	Pred No. 2	2e-16	
Matches	311	Conservative	0	Mismatches	231
				Indels	27
				Gaps	2
Qy	55	CTCCCTGGCGGGGAGAAAGCACTTATACGCGCTCTGTGTCTCCGAGGTCTTTTGAG	114		
Db	1137	CTGCCCTTGGCAACAGGGCATCAACCCCTTACCGGGTGTGGTCTCGAAATCATGTGTGAG	1078		
Qy	115	CAGACCCGGGTGAGCAGGCGCTCTCCCTTATACCCGCTTCTGAGGCGTTTCCACC	174		
Db	1077	CAGACCCAGGTGAGCAGCGTGTCTGGTTACTTGCACGTTTCATGCGAGCGTTGCCAG	1018		
Qy	175	CTGAAGGCTCTGGCGCGGCTTCTCTGGAAGAGTCTTAAAGGTCTGCAAGGGGGGGG	234		
Db	1017	GTCGAGGCACTGGCCCGGGCGGCGGAGAGCAAGTCTGCACCTGTGAGCCGGGCTGGG	958		
Qy	235	TACTACCGCGGAGCGGAAACCTCAACCGCTGTGCGCCGAAGCGT-----G	279		
Db	957	TACTACAGCGTGTGCGGCAACTGACACAAGACCGCGCAGATGATGTGTGAGCGGATCG	898		
Qy	280	GAGAGACTTCCCCCGAGCTTTCGCCAGCTTTCGGGGGCTTCTGTGTCTGGGCTTACAAC	339		
Db	897	GGGGAATTCCTCCCGGACGTCGTAGCAACTCGCGGAATCGCCGGGCAATCGGCGCTCAC	838		
Qy	340	GCGGCGGCGGTGACTTCATCGCTTTCGGGAGACGGGTGCGAGCGGTGACCGGAACGTC	399		
Db	837	GCTGTGAGGCATGCGCAGCTGTGTGATGTGGCTGTGCGGCAACGATCTCTGACGGCAACGTC	778		
Qy	400	CGAGGGGTCTTCTCCGCGCTTTCGCCCGGAGAAAGCCCAAGAAAGAGGCTTTTGCG	459		
Db	777	AACGCGTATGTGGCGGCTACTACGTGGGCAAGAGGGGTATTCGGGCGAACCGAAGGTGCC	718		
Qy	460	CTGGCCAGGGCTCTCTCCCGAGGGCGTGAACCCCGGGGTGTGG-----AAG	507		

Db 717 AGGGCGCTGTGGGAAAGCGCGGAAAGCTTTCACCCCGCACGCAAGGGTCAACCACTACACC 5658  
OY 508 CAGACCCCTATGTAGACTGGGGGCGACAGGCTGTGCGCTGTGCGAAAGGGCCGCTGGCGGGCC 5657  
Db 657 CAGGCGATATGACCTCGGGCCGACACCTCTGTGCAAGCGCAGCAAGGCCAATGGCTGTCTT 598  
OY 568 TGGCCCTCTAGGGGCTTTCGCGGGGGAA 596  
Db 597 TGCCTCGTGTCTCGGCTGCGCGCGCA 569

```

1      RESULT 3
2      US-09-252-991A-6739
3      / Sequence 6739 Application US/09252991A
4      / Patent No. 6551795
5      / GENERAL INFORMATION:
6      / APPLICANT: Marc J. Rubenfield et al.
7      / TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
8      / TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
9      / FILE REFERENCE: 107196.136
10     / CURRENT APPLICATION NUMBER: US/09/252,991A
11     / CURRENT FILING DATE: 1999-02-18
12     / PRIOR APPLICATION NUMBER: US 60/074,788
13     / PRIOR FILING DATE: 1998-02-18
14     / PRIOR APPLICATION NUMBER: US 60/094,190
15     / PRIOR FILING DATE: 1998-07-27
16     / NUMBER OF SEQ ID NOS: 33142
17     / SEQ ID NO 6739
18     / LENGTH: 1413
19     / TYPE: DNA
20     / ORGANISM: Pseudomonas aeruginosa
21     / US-09-252-991A-6739

```

Query Match	Similarity	12.9%	Score 125.4	DB 4	Length 1413
Best Local	Similarity	54.7%	Pred. No. 2.2e-16		
Matches	Conservative	0	Mismatches 231	Indels 27	Gaps 2
QY	55	CTCCCTGGCGGGGAGAGAGACCTCTTACCGCGTCTGTCTCGAGTCTCTTGAG	114		
Db	409	CTGCCCTGGCAGCAGAGGACATACCCCTTACCGGGGTGGGTCTCGAATCATGCTGCAG	468		
QY	115	CAGACCCGGGTGGAGCAGACCCCTCCCTTATTAACGCGCTTCTTGAAGCCTTTCACAC	174		
Db	469	CAGACCCAGGTGAGCAGCACCGTGTCTCGTTACTTTCGACCGTTTCATGACAGCTTTCACAC	528		
QY	175	CTGAAGCGCCCTGGCCGCGCGCTTCCCTGGAGAGGATCTTATGGGTCTGCAGGGGCGCGGC	234		
Db	529	GTGAGGCACTGGCCGCGCGCGCGCGAGAGCAATGTTCTTGCACTGTGACCGGCGCTCGGC	588		
QY	235	TACTACCGGCGGCGGAGAACCTCTTCACCGCTGGCCCGAAGCT-----G	279		
Db	589	TACTACAGCGGTGGCGCGCACTGCACAAGACCGCGCAGATGTGTGTGACGCGCATGGC	648		
QY	280	GAGGAGCTTCCCCCGAGCTTGGCCGAGCTTGGGGGGCTTCTGTGTCTGGGGCTTAAAC	339		
Db	649	GGGGAGTTCCCCCGGACCTTCAGAGCACTCGCGCAATCGCCGGCAGTCGCGCTTCAC	708		
QY	340	GGGGGGGGGTGGGCTCCATGCTCTTGGGGAGAGCGGTGGCGGCGGTGGACGGAAAGTC	399		
Db	709	GCTGAGGCATGACCCAGCTGTGTGATGGCTGTGGCGACGACGATCTCGACGGCAACGT	768		
QY	400	CGGAGGTCCTCTCCCGCTCTTTCGCGCGGGAAGCCCAAGAGAGAGGCTTTTGC	459		
Db	769	AAGCGGTACTGGCGCGCTACTCGGGGCAAGACGGCTATCCGGCGAACCGAAGGTGGC	828		
QY	460	CTGCGCCAGGCTCTCTCCCGAGGGCGTGAACCCGGGGGTGTG-----AAC	507		
Db	829	AGGCGCTGTGGAGAACCCCGCAACGCTTCAACCCGCAAGCAGGTCAACCACTACAC	888		
QY	508	CAGGCGCTATGAGAGCTGGGGGCAACGATCTGCTGCGGAAACGCGCGCTTGGGGGCGC	567		
Db	889	CAGCGCATATGAGACTTGAGCGCACTCTTGTGACGCGCAGCAAGCCAGTTGCTGCTT	948		

QY 568 TGCCCCCTAGGGGCTTTCTGCGGGGAA 596  
DB 949 TGCCCCGTGTCTCCGCTGCGCGCA 977

RESULT 4  
US-09-489-039A-4788  
Sequence 4788, Application US/09489039A  
Patent No. 6610836

GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
PRIOR FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 4788  
LENGTH: 1551  
TYPE: DNA  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-4788

Query Match 11.6%; Score 113; DB 4; Length 1551;  
Best Local Similarity 53.0%; Pred. No. 6e-14;  
Matches 307; Conservative 0; Mismatches 245; Indels 27; Gaps 2;

QY 53 CCTCCCTGCGGGGAGAGAACCTTACCGGCTCTGAGGCTTCTGCG 112  
DB 560 CCTCCCTGCGCAATGCGCAAGACGCTTCAAAATGCTCTCGAATGATGTCG 619  
QY 113 AGGAGACCGGGGTGAGAGAGCCCTTACCGGCTCTGAGGCTTCTGCG 172  
DB 620 AACAAACCGAGATACCAAGGTAATCCCTTATTGAAAGCTTTATGCGGCTTCCCA 679  
QY 173 CCTGAAAGGCTGCGCGGCTTCCCTGAGAGGCTCTTAAAGGCTCTGAGGGGGCG 232  
DB 680 CGGAGGATCTGCGCAAGCGCGGCTGATGAGATCTGATCTGAGACGGTCTGG 739  
QY 233 GCTACTACCGGCGGCGGAACAC-----CTCACCGCTGCGCGGAGCG 277  
DB 740 GCTACTACCGGCGGCGGCACTTACATAAAGCGCGCAAGTCCACACTGACG 799  
QY 278 TGGAGAGCTTCCCGGAGCTTCCGCGGCTTCTGAGGCTTCTGAGGCTTACA 337  
DB 800 GCGGGAATTCCTCCGAGCTTGAAGAGTGGCGGCTTACCGGCGTGGCGCTCA 859  
QY 338 CCGGCGCGGAGCTCTCAATGCTTCTGAGAGGCGGAGTGGCGGCTGAGCGGAA 397  
DB 860 CGGCGGGGCGATTTTATCTCTTCTGCTGAGTCAAGTATTCGATTCGAGCGCA 919  
QY 398 TCCGAGGAGCTCTCTCCGCTCTTCTGCGCGGAAAGCCCAAGAGAGAGCTTTTCG 457  
DB 920 TGAAGCGGCTGCTCGCGCTGCTATGCTGACAGCGCTGGCGGGAAGAGAGTGG 979  
QY 458 CCTCGCCCAAGGCTCTCCCGCGGAGGAGTGAAGCGGAGGAGTGG-----A 505  
DB 980 AAAAAGCGCTGAGAGATCAAGAGAGTCAAGCGGAGAGGCGCTGAGCGCTTCA 1039  
QY 506 ACAAGGCTCTATGAGCTGAGGAGCAAGTCTGCTGCGAAAGCGCGCTTGGCGGG 565  
DB 1040 ACAAGGAGATGATGATCTCGGAGCAAGTGTGCAAGCGGCGAAGTGCAGC 1099  
QY 566 CTTGCGGCTTCTGAGGCTTCTGCGCGGAGAGAGAGCGCC 604  
DB 1100 TGTGTGCTGAGCAAGCGCTGCTGAGCGCTACGCGAAC 1138

RESULT 5  
US-09-252-991A-6783  
Sequence 6783, Application US/09252991A

Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6783  
LENGTH: 549  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6783

Query Match 11.0%; Score 107.2; DB 4; Length 549;  
Best Local Similarity 54.4%; Pred. No. 7.5e-13;  
Matches 248; Conservative 0; Mismatches 193; Indels 15; Gaps 1;

QY 55 CTCCCTGCGGGGAGAGAACCTTACCGGCTCTGAGGCTTCTGCGAG 114  
DB 60 CTGCTGCGGAGAGAGGAGATCAAGCTTACCGGCTGAGGCTTCTGCGAG 119  
QY 115 CAGACCGGAGTGAAGAGGCTTCCCTATTAACCGGCTTCTGAGCGCTTCCAC 174  
DB 120 CAGACCGAGTCAAGACCGGCTGAGTCTGAGCGGTTTCAAGCGAGCGGCTGAG 179  
QY 175 CTGAGGCTTCTGCGGCTTCTGAGAGGCTTCTGAGGCTTCTGAGGCTTCTGAG 234  
DB 180 GTGAGGCTTCTGCGGCTTCTGAGAGGCTTCTGAGGCTTCTGAGGCTTCTGAG 239  
QY 235 TACTACCGGAGGAGAGAACCTTCAAGCGGCTGAGCGGAGGCT-----G 279  
DB 240 TACTACCGGAGGAGAGAACCTTCAAGCGGCTGAGCGGAGGCTTCTGAGGCTTCA 299  
QY 280 GAGAGCTTCTCCCGAGCTTCCGAGCTTCTGAGGCTTCTGAGGCTTCTGAGGCTTCA 339  
DB 300 GAGAGCTTCTCCCGAGCTTCCGAGCTTCTGAGGCTTCTGAGGCTTCTGAGGCTTCA 359  
QY 340 GCGGCGGAGTCTCTCAATGCTTCTGAGGAGGCTTCTGAGGAGGCTTCTGAGGAGT 399  
DB 360 GCTGAGCAATGCGCAAGCTTCTGAGGAGGCTTCTGAGGAGGCTTCTGAGGAGT 419  
QY 400 CGGAGGCTTCTCCCGCTTCTGAGGAGGCTTCTGAGGAGGCTTCTGAGGAGT 459  
DB 420 AAGCGGCTTCTGAGGAGGCTTCTGAGGAGGCTTCTGAGGAGGCTTCTGAGGAGT 479  
QY 460 CTGCGCCAGGCTTCTCCCGAGGCTTCTGAGGAGGCTTCTGAGGAGT 495  
DB 480 AGGCGCTTCTGAGGAGGCTTCTGAGGAGGCTTCTGAGGAGT 515

RESULT 6  
US-08-813-574-1  
Sequence 1, Application US/08813574  
Patent No. 6013473  
GENERAL INFORMATION:  
APPLICANT: Wei, Ying-Fei  
TITLE OF INVENTION: Human Muty  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

```

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/813.574
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/013,132
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1858 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-813-574-1

```

```

Query Match      10.7%; Score 104; DB 3; Length 1858;
Best Local Similarity 54.0%; Pred. No. 3,6e-12;
Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;

QY 69 GAGAGAGACCTTACCGCGCTGCTGCTCCGAGTCTCTTGACAGACCCGGGTGGA 128
Db 498 GAGACAGGGGAGATATGCTGTGTGTCTCAGAGATCATGTGTGACAGACAGGTTC 557
QY 129 GCAGAGCCCTCCCTATTACCGCCGCTTTTGAAGCGCTTCCACCTGAAAGCGCTGC 188
Db 558 CACTGTATCAACTACTATACCGATGATGACAGATGCGCTTACCTGAGAGACCTGAC 617
QY 189 CGCGGCTTCCCTGGAAGAGCTTTAGGCTCTGCGAGGGGGCGGCTTACCGGGCGGC 248
Db 618 CAGTGTCTTCCCTGAGAGAGTGAATCACTGTGGGCTGGGCTGAGCTTATTTCTGTGG 677
QY 249 GGAACACTCTCAACCGCTGCGCGAGCGTGAAGAGCTTCCCGGAGCTTTC----- 300
Db 678 CCGCGGCTGCGAGAGGAGAGCTCGAAGGTGTGTAGAGACTAGGGGGCCACATGCGACG 737
QY 301 -----GCGAGCTTGGGGGCTTCTGTGTCTGGGCTTACACCGCGCGCGGT 350
Db 738 TACAGCAGAGACCTTGAGAGAGCTCTGCTGTGGGTGGGGCGCTTACACAGCTGGGGCAT 797
QY 351 GGCCTTCATGCGCTTCCGGAGAGCGGAGTGGCGGCTGAGCGGAGACGTCCGAGGGTCT 410
Db 798 TGCCTTATGCGCTTGGCGAGGAGACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 857
QY 411 CTCCCGCTCTTGGCCC--GGAAGAGCCCAAGAGAGAGAGCTTTTGGCCCTGCGCCAG 468
Db 858 GTGCGCGTCTCGAGCCATTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 917
QY 469 GGCCTCTCTCCCGAGGAGCTGAGACCG-----GGGGTGTGAGACAGGCGCTCAT 518
Db 918 GGGCTTAGCCACAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 977
QY 519 GAGAGCTGGGGGAGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 578
Db 978 GAGAGCTAGGGGAGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1037
QY 579 GGCCTTCTGCGGGG 592
Db 1038 GAGCCTGTGTGCGGG 1051

```

```

US-09-651-656-38
; Sequence 38, Application US/09651656
; Patent No. 6340566
; GENERAL INFORMATION:
; APPLICANT: MCCUTHEN-MALONEY, SANDRA
; TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE
; TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,
; FILE REFERENCE: IL-10689
; CURRENT APPLICATION NUMBER: US/09/651,656
; CURRENT FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/192,764
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 2293
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-651-656-38

```

```

Query Match      8.9%; Score 87.2; DB 4; Length 2293;
Best Local Similarity 54.1%; Pred. No. 7,2e-09;
Matches 210; Conservative 0; Mismatches 163; Indels 15; Gaps 1;

QY 53 CCTTCCCTGCGGGGAGAGACCTTACCGCGCTGCTGCTCCGAGTCTTCTGTC 112
Db 1161 CTCTGCGCTGAGAAATTGACAGACGCCCTCAAGATATGCTCTCAGAGTATGTTGC 1220
QY 113 AGAGAGCCGGGTGAGAGAGCGCCCTTCTTACCGCGCTTCTGAGAGCGCTTCCCA 172
Db 1221 AACAACTAGATGCTGACCGCTTATCCCTTATTGAACGCTTATGTGGCGCTTCCCGA 1280
QY 173 CCTGAGGCGCTGCGCGGCTTCTCTGGAAGAGTCTTATGAGTCTGCGAGGGCGG 232
Db 1281 CGGTGACGATCTGCGCATGTGCGCGTGAAGAGTCTCCACTGTGTGTGTGTGTGTGTGT 1340
QY 233 GCTACTACCGGGGCGAAGACCTTCAACCGCTGCGCGGCGGAGCGGTGAG----- 282
Db 1341 GCTATTACCGCGCGCGCGCATGTGATTAAGCGGCGCAACAGTGTGCGACTTACACG 1400
QY 283 -----GACCTTCCCGAGCTTCTGCGAGCTTCTGCGGGCTTCTGCTGCGGCTTACA 337
Db 1401 GCGGTAATTTCCGGAACCTTTGAGGAAGTGTGACAGACTCGGCGGTGCGGCTTCCA 1460
QY 338 CCGCGCGCGGCTGCTTCAATCGCTTCTGCGAGCGGGTGTGCGCGGTGTGACCGGAACG 397
Db 1461 CCGAGGGGCGATCTCTGCTTCTGCGGTGTGACACTTCCGATTCTGACGCGTAACG 1520
QY 398 TCCGAGGGTCTCTCCCGCGCTTTCG 425
Db 1521 TCMAACGCGTGTGCGCGCTGTATGC 1548

```

```

RESULT 8
US-09-650-855-38
; Sequence 38, Application US/09650855
; Patent No. 6365355
; GENERAL INFORMATION:
; APPLICANT: MCCUTHEN-MALONEY, SANDRA
; TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE
; TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA
; TITLE OF INVENTION: MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA
; FILE REFERENCE: IL-10284
; CURRENT APPLICATION NUMBER: US/09/650,855
; CURRENT FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/192,764
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38

```

```
; LENGTH: 2293
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-650-855-38
```

Query Match 8.9%; Score 87.2; DB 4; Length 2393;  
 Best Local Similarity 54.1%; Pred. No. 7.2e+09;  
 Matches 210; Conservative 0; Mismatches 163; Indels 15; Gaps 1;

Qy	5	CCCTCCCTGGCGGGGAGAAAGACCCCTTACCGGGTCCGTGCTCCGAGGCTCTTCGCG	112
Db	1161	CTGTGCCCTGGCAAAATTGACAAAGAGCCCTTACAAAGTATGCGCTCTCAGAGATGATGTTGC	1220
Qy	113	AGCAGACCCGGGTGAGCAGGCGCCTCCCTTATACCGCGCTTTCTGAGGCGCTTCCCA	172
Db	1221	AACAAACTCAGGTGTGCGACCGTTATCCCTTATTTGAAACGCTTATATGGCGGCTTCCGA	1280
Qy	173	CCCTGAAAGCCCTTGCGCGCGGCTTCCCTGGAAGAGTCTTAAAGGTCTGTGACAGGGGGCG	232
Db	1281	CGGTAGACCGAATCTGCGCCATGCGCGCGCTGACGAAAGTCTCACTGTGAGACCGGGCTTG	1340
Qy	233	GCTACTACCGGGGGGCGGACACCTCCACCGGCTGGCGCCGGAAGCGTGGAG-----	282
Db	1341	GCTATTACGCCCGCGCGCGCAATCTGCATTAAGCGGACCAACAGGTGCGACTTTACACG	1400
Qy	283	-----GAGCTTCCCGCCGAGCTTTCGCGGAGCTTTCGGGGGCTTCTGTGTCGGGCTTACA	337
Db	1401	GCGGTAAATTTCCCGGAACCTTTTGAAGAAATTGACGACATCGCGGGCGTCCGGGCGTTTCA	1466
Qy	338	CCGCGGCGGCGGTGCGCTTCATTCGCTTTCGGGGAAGCGGGGTGCGCGGCTGACCGGAACG	397
Db	1461	CCGCGAGGCGCGAATCTCTTCGCTTTCCTGTGGGTAAGACCTTTCGAAATCTTCAGACGGTAAAG	1520
Qy	398	TCGGAAGGCTCCTCTCCGCGCTCTTCGCG	425
Db	1521	TCAAACGCGTGTGCGCGCGCTGCTAAATGC	1548

RESULT 9  
US-09-103-840A-2  
; Sequence 2, Application US/09103840A

```

? GENERATED INFORMATION:
? APPLICANT: FLEISCHMAN, Robert D.
? APPLICANT: WHITE, Owen R.
? APPLICANT: FRASER, Claire M.
? APPLICANT: VENTER, John C.
? TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
? TITLE OF INVENTION: TUBERCULOSIS
? FILE REFERENCE: 24366-20007.00
? CURRENT APPLICATION NUMBER: US/09/103,840A
? CURRENT FILING DATE: 1998-06-24
? NUMBER OF SEQ ID NOS: 2
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 2
? LENGTH: 4403765
? TYPE: DNA
? ORGANISM: Mycobacterium tuberculosis
? FEATURE:
? OTHER INFORMATION: CDC 1551
? OTHER INFORMATION: "n" bases at various positions throughout the sequence
? OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

	Query Match	Score	DB 3	Length
Best Local Similarity	50.2%	Pred. 1.18e-08		4403765
Matches	345	Conservative	0	Mismatches 309; Indels 33; Gaps 4

  

QY	22	CTTCTCGGCTGTATCCGCGAAAGCCCGCCCCCTCCGCTGAGCGGCGG---GAGAGGAC	78
Db	4022631	CTTCTCGCTTGTATCAGAGATCGCAGCCGGGATCTGCCCCGTGCGAGAGCCCGCGTGTACGC	4022690

  

QY	79	CCTTACCGCGCTCTGCTCTCCGAGGCTCTTCTGACAGACACCCGGGCTGAGACAGCCCTTC	138
----	----	---	-----

4022691 CCGTGGCAGATCTCTGGTCAAGCGAGTTCAATGCTGACAGAGACCGCGGCGCGCCGGGTGCTG 4022750  
4022751 GCATCTGGCGCGGACCTGGAGCGCGCGGTGGCCACACCGCTGGCCACCGCCACAGGGCAGC 4022810  
4022811 ACCGCCAATGTTGTTACCGCGCTGGGCGAAGCTGGGTATCCAGGGCGAAGCAAGCGCTTA 4022870  
4022871 CCGGAGAGAGTCCCTTAAAGGTCTGGACAGGGGCGGGCTACTACCGCGGGCGGAGAACCTC 258  
4022871 CACCAAGTGCACACCGTCACTGCGCCCGGACACAAATACGTGGTGCCTCCAGATTCGAG 4022930  
4022931 ATCTGGTCAACCTGCGCGGCGGTGGAGGTACACCGCGCGCGGGTGGCTGTTGCT 4022990  
4022991 TTGGGAGAGCGGGGTGCGCGGTGACGGGAAAGTCCGGAGGTTCTTCCCGCTCTT- 422  
4022991 TACCGCAGCGGGGTGCGGTGAGACCAAAATGGTGGCGCGGTGGCGCGCTT 4023050  
4023051 CACGGCGCGCGCCAGCGCGGTGCGCCATCGGTGCGCGGACACACGCGCAGATTTGGCG 4023110  
4023111 CTGTTGGCGGACCGCGAGACGGCCCTGAAATTTTCGGTTCGCGTACGATGAGTTGGTGGC 4023170  
4023171 ACGGTCTGCGCGGAGAAAGCGCCCGTTCGGCGGGCGCTGCCTTACGGGAGCCTTCTGCGG 591  
4023231 CGGATGCGGGTTATTCGCGCGTTCGAGAGGTCC--GCCGCGCGCGGGGCAAGGCTTACAC 4023290  
4023291 CTGCTGCGCGCTGCTCTCTCTGCGGCGG 678  
4023291 GGAACCGACCGCCAAAGTCCGCGACGG 4023314

```

US-09-103-840A-1
RESULT 10
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRA
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H7RV
US-09-103-840A-1

```

```

Query Match      8.9%;   Score 86.6;   DB 3;   Length 4411529;
Best Local Similarity 50.2%;   Pred. No. 1.8e-08;
Matches 345;   Conservative 0;   Mismatches 309;   Indels 33;   Gaps 4;

```

0y  
22 CTCTCGCCTGAGTACCCGGGAAAAAGCCCCCCCCTTCCTGTGCAGGGG---GAGAAGGAC 78

```

Db 403050 CTTCGCTTGTATGAGGATGCAACCGGATCTGCGGAGAGCCCGGTCTCAGC 4030609
Qy 79 CCTTACCGGCTCTTGTCTTCCAGGTCCTTCTGACAGACCCGGGTGAGAGCCCTC 138
Db 4030610 CCGTGGAGATCTGTGTACAGCGATTCATGCTGACAGACCCGGCGCCGGGTGCTG 4030669
Qy 139 CCCTATTACCGCGCTTTCTGAGCGCTTTCCACCTGAAAGCCCTGCGCCGCTTCC 198
Db 4030670 GCGATCTGGCGGACTGGGTGCGGCGGTGGCCACCCCTGCGCCACCGCCAGC 4030729
Qy 199 CTGGAAGAGGCTCTTATGAGTCTGAGCAAGGGGCGGCTACTACCGCGGCGAGACCTC 258
Db 4030730 ACCGCGGATGTGTACGCGCTGAGGCAAGCTGGGCTATCCAGGCGAGCCAGGCTTA 4030789
Qy 259 CACCGCGCTGCGGAGAGCGTG-----GAGAGCTTCCCGAGCTTCCGCC 303
Db 4030790 CACGAGTGGCCACCGCTCATTGCGCCCGGACCAATGACGTGTGCTCCGAGATATCGAG 4030849
Qy 304 GAGCTTGGGGGCTTCTGTGTCTGAGCTTACACCGCGCGCGCGCTGAGCTTCATCGCC 363
Db 4030850 ATCTGTGTACCTCTGCGGCGCTGAGAGCTACACCGCGCGCGGTGCTGCTTCTGCT 4030909
Qy 364 TTGCGGGAGCGGCGCGCGCTGAGCAAGGAACTGCGAGGCTCTCTCCGCTCTT- 422
Db 4030910 TACCGCCAGCGGCTGCGGTGTGAGCAACATGTGCGCGCGGTGTGCGCGCGCTT 4030969
Qy 423 -----CGCCGGGAAAGCCCAAGAGAGAGCTTTTCCGCTGCGCCAGGCGC 471
Db 4030970 CACGCGCGCGCGAGCGCGGTGCGCATGCGTGCCTGCGCCGACACGCGAGCTTTGGCG 4031029
Qy 472 CTCTCCCGGAGGCGGTGAGACCGGGGCTGTGAAACAGGCGCTCATGAGACTGCGGCGC 531
Db 4031030 CTGTGCGGACCGGAGACCGCGCTGAAATTTTGTGCGCGCTGATGAGATTGGGTGCG 4031089
Qy 532 ACGGTCTGCTGCGGAAAGCGCGCGGTGCGGCGCGCTGCGCGCGCTTCTGCGCG 591
Db 4031090 ACGGTGTGACCGCGCGCGCGACACCGCGGTGTTATGCGCGTGTGAGCTGTGCGAGTGG 4031149
Qy 592 GGGAAAGAGCGCGCGCGGCTGTACCGCGCGCGCGCGAGAGCGCGCGAGAGAGCGCG 651
Db 4031150 CGGATGCGCGGTATTCGCGCGCTCGAGCGGTCC--GCCGCGCGCGGGGCGCTACAC 4031206
Qy 652 CTGCTGCGCGCTGTCTCTCTGCGGCGG 678
Db 4031207 GGAACCGACCGCGCAAGTCCGCGAGCG 4031233

```

```

RESULT 11
US-08-961-527-12/c
; Sequence 12, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunesh
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9909 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-12

```

```

Query Match      8.4%; Score 82; DB 4; Length 9909;
Best Local Similarity 52.2%; Pred. No. 8.5e-08;
Matches 218; Conservative 0; Mismatches 185; Indels 15; Gaps 1;

```

```

Qy 22 CTCTGCGCTGTACCGGAAACGCCCGCCCTCCCTGCGCGGGGAGAGAGCCCT 81
Db 5808 CTCTGCGCTGTATGATGAAACAAAGAGATTGCTTGAGAGAGAAATTAATCT 5749
Qy 82 TACCGCGCTGTGTCTCCGAGTCTTTGCAAGACCGGCTGAGAGAGCCCTCCC 141
Db 5748 TATCATCTGTGGTATCTGAATCATGCTTCAGAGACAGGGGTGATACAGTTATCCT 5689
Qy 142 TATTACCGCGCTTCTGAGAGCTTTCACACCGCGAGCGCTGCGCGGCTTCCCTG 201
Db 5688 TACTACGAAAGATTTTTGACTGCTGTTTCAACTGTGCAAAAGTCTGGCACTGGCGCTGAG 5629
Qy 202 GAAAGAGCTCTTGAAGGTGTGCGAGGGGCGGCTA-----CTACCGCGCG 246
Db 5628 GAGAGTTTACGAAAGCTTGGAGGGCTTGCGCTATATTCTCGAGTTGCAATATGCGAG 5569
Qy 247 GCGGAACACTTCAACCGCGTGGCCGGAAGGTGAGAGAGCTTCCCGAGCTTCCCGAG 306
Db 5568 GCTGAGCGCCAGCAATTAATGACTGACTTGTGTGCGCAATTTCCAAATACCTATGAGAG 5509
Qy 307 CTGCGGGGCTTCTGTGCTGCGGCTTACACCGCGCGCGGTGAGCTTCATGCGCTTC 366
Db 5508 ATTTCAGCTTGAAGAGGATTTGACCTTACAGAGAGAGCCATTTCAGATTTGCTTTT 5449
Qy 367 GGGAGGGGCTTCTGTGCTGCGGCTTACACCGCGCGCGGTGAGCTTCCTCTCCGCTTTG 424
Db 5448 AACTTGCTGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 5391

```

```

RESULT 12
US-09-107-532A-3654
; Sequence 3654, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID and AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998

```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arinello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 893-5007  
TELEFAX: (781) 893-8277  
INFORMATION FOR SEQ ID NO: 3654:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1212 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...1212  
SEQUENCE DESCRIPTION: SEQ ID NO: 3654:  
US-09-107-532A-3654

Query Match  
Best Local Similarity 47.6%; Pred. No. 0.00029; Indels 27; Gaps 2;  
Matches 278; Conservative 0; Mismatches 279;

31 TGGTACCGGAAAAAGCCGCCCTCCCTGCGCGGGGAGAAAGACCTTACCGCGTC 90  
100 TGGTATACAGAAAAAGAAACCTCGCTTGGCGCTATATCGTATCTTATCGTATC 159  
91 CTGCTCCGAGAGTCTCTTGCAGCAGACCGGGGTGAGACGCGCTCCCTTATACCGC 150  
160 TGGATCTCTGAGATCATGCTTTCAGCAACAGAGTACACAGATGATGATTTAT 219  
151 CGCTTTTGAGAGCCTTCCACCTGAGGCGCTGCGCGGCTTCCCTGAGAGAGTC 210  
220 CGTTTCAATGAAATGTTCTTCAATGAGAGCTTACAAATGCTTCAAGAAAGAACTT 279  
211 CTGAGGCTTGCAGGGGCGGCTTACTACCGCGCGGAAACACTC-----258  
280 TTTAAACATGAGGAGGCGCTTGTATTAATCAAGAGCGCGAATCCAGCTGCTG 339  
259 ---CACCCTGAGCCGAAAGGTGAGAGAGCTTCCCGAGCTTCCCGAGCTTCCGAGG 315  
340 AATAGATCATGTGAGATTTGATGGGAAATGCGCAACGCTGAAGAAATTAAGTCA 399  
316 CTTCCTGCTCGGCGCTTACACCGCGCGGCGGTGCTCCATCGCTTCCGAGAGCG 375  
400 TTGAAGGAAATGAGACATATACAGAGCGCATGATGAAATTCGACTTTGACCTTCA 459  
376 GTGCGGCGGTGAGACGGAACGTCGAGAGGCTCTTCCCGCTCTTTCGCGCGGAAAGC 435  
460 GAACCTGAGTTGACGCGACGTGATGAGATGACAGATGTTGTTTGAATGAAACA 519  
436 CCCAAGAGAA-----GAGCTTTTCCCGCTCGCGCAAGGCGCTTCCCGAG 483  
520 GACATCGAAGAAAGCTTACAGAGAAATCTTTGATGAAAGGATGCGGAAATCATGAT 579  
484 GCGGTGACCGCGGGGTGTGAACAGGCGCTCATGAGCTTCGGGCGCAAGCTGCGCTG 543  
580 GAAAGCATTCAGATGAAATTAACCAAGCAATGATGATGATGATGATGATGATGATG 639  
544 CCGAAACGCGCGCGCTTGCAGGCGCTTCCCTTACGAGGCGCTTCTG 587  
640 CCAACTTCTCCAAAGTGTGAGATTTGTCGATCCAGGCTTTTG 683

RESULT 13  
US-09-134-000C-1710  
Sequence 1710, Application US/09134000C  
Patent No. 6617156  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 032796-032  
CURRENT APPLICATION NUMBER: US/09/134,000C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/055,778  
PRIOR FILING DATE: 1997-08-15  
NUMBER OF SEQ ID NOS: 6812  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1710  
LENGTH: 1188  
TYPE: DNA  
ORGANISM: Enterococcus faecalis  
US-09-134-000C-1710

Query Match  
Best Local Similarity 48.8%; Pred. No. 0.0005; Indels 15; Gaps 1;  
Matches 211; Conservative 0; Mismatches 206;

8 CTTGCGGAAAGCCCTTCTCGCTGTACCGGAAAGCCCGCCCTCCCTGCGCGG 67  
53 CATTCCAGAGAGATTTTATGCTGTATGAGACGAGAAAAAGCAATTTACCTTGGCGAG 112  
68 GGGAGAGAGACCTTACCGCGCTGCTCCAGGCTCTTGCAGACAGACCGGGGTG 127  
113 CGAATACAGTATGATTTTATGATTTATGATTTATGATTTATGATTTATGATTTATG 172  
128 AGCAGGCGCTTCCCTTATTAACCGCGCTTCTGAGAGCGCTTCCACCTGAGAGCGCTG 187  
173 ATACATGATGATTTATTTTATGATTTATGATTTATGATTTATGATTTATGATTTATG 232  
188 CCGCGGCTTCCCTGAGAGAGTCTTAAAGGCTCTGAGAGGCGGGGCTACTACCGCGG 247  
233 CGAAGGCGCAGATGATGATTTGATTAAGCTTGGAGAGGTTAGTTACTATTTACAGAG 292  
248 C-----GAAACACTCCACCGCTGCGCGGAGAGAGCTTCCCG 292  
293 CGGCTAATTTAAAGTGGACGCGCAACAGATTTTTCAGAAATTTGGTGGAAAAATGCTG 352  
293 CGAGCTTCCGAGCTTCCGAGGCTTCTGCTCGGCGCTTACACCGCGCGGCGGTG 352  
353 ACACATCGAAGATTTTCGAGTTTAAAGAAATCGGCCCTTACGCGCTGTCGATG 412  
353 CTTCATCGCTTCCGAGAGCGGCTGCGGCGGTGAGACGGAACGTTCCGAGAGGTCCTCT 412  
413 GCAGTATTTGCTTAAATCTCCGGAACAGCAATTAACGATGATGCGGAGTGA 472  
413 CCGCGCTCTTGG 424  
473 GCGGTTATTTG 484

RESULT 14  
US-09-385-028-23  
Sequence 23, Application US/09385028  
Patent No. 6232106  
GENERAL INFORMATION:  
APPLICANT: Susan E. Jensen  
APPLICANT: Kwame A. Aido  
APPLICANT: Ashish S. Parthasarathy  
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic  
Patent No. 6232106  
TITLE OF INVENTION: Acid Biosynthesis  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, P.L.C.



```

STREET: The Jennifer Building, 400 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,028
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/790,462
FILING DATE: 29-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: D. Douglas Price
REGISTRATION NUMBER: 24,514
REFERENCE/DOCKET NUMBER: 1418/P57452US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-6666
TELEFAX: (202) 39305350
TELEX: RCA 248593 IDEA UR
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1227 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-385-028-23

```

```

Query Match      6.4%; Score 62.2; DB 3; Length 1227;
Best Local Similarity 44.4%; Pred. No. 0.00055;
Matches 345; Conservative 0; Mismatches 423; Indels 9; Gaps 2;

```

```

157 CTGGAGCGCTTCCACCTGAGAGCCCTGCGCGCTTCCCTGAGAGAGTCTTAGG 216
88 CTGGCGGGGCTGCGGTCCAGAGAGCGCGAGCGGGTGAAGCTGTGGGAGCGGAGCAG 147
217 GTCTGGAGAGGGGGGGGCTATACCGGCGGGGGAACAACCTCCACCGCGCTGGCCGAGC 276
148 GTGTGGCTGTGAGACTCGACGCGGGAGCCCGGCGCTTCTGGGCGAACCGCGCTTAC 207
277 GTGAGAGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTAC 336
208 GCGGTGACGAGCGCGCGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTAC 267
337 ACCGCGGGGGGGGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTAC 392
268 AACCGGAGTGGGGTCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTAC 327
393 --GAACGTCGCGAGGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTAC 450
328 ATGTACACCGGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTAC 387
451 CTTTTCGCTCGCGAGGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTAC 510
388 CTCGTGACGAGATCTGGGGGGGGTGTGAAGGGGAGCGGGCTGTGACCTGTGCGCC 447
511 GCGCTCATGAGAGCTGGGGGGGCAAGGTCTGCTGCGCGAGAACCGCCCGCTTGGGGGGGCTGCG 570
448 GAGCTGAGAGATCCCGGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTAC 507
571 CCCCTAGAGGCTTCTGCGCGGGAGAGAGCGCCCGGCGCTACCCCGCGCCGAGAG 630
508 GCGCGGAGGTATGAGAGAGAGCGCGCGGTCTCTCATGACCGCGCTTACCGCGAG 567
631 CGCCGGGAGAGAGAGAGCGCTGCTGCGCTTCTGCTGCTGCGGGGAGAGGGGTGAC 690
568 CAGGTGCGCAAGGCGCGGAGAGAGCTGACGAGCTTCTGCGGGAGAGCTGTGAGAGAGCG 627

```

```

691 CTGAGAGAGCTTGAAGGGGCGCTTCAAGGCTCTAAGCGCTTCCCTCTTCCCTGAG 750
628 ATCAGAGAGCGGGGAGCGAGCTGATGAGCGCGCTCTCATGACAGAGGAGCGCGGG 687
751 GAGCTTCCCGGGGGAGGGGGCTTGGGGTGAAGCTTGAAGCGCTTGAAGGAGTGGC 810
688 CATCTGCGGGTGAAGAGATGATCCCATGTGCGGCTGCTGCTGCTGCGCGAGCGC 747
811 CAGCCCTCAACCAAGAGGCTTGGGTGAAGGTGCGGGGGCGCTTGGAGAGGAG 870
748 ACACACCAAGAGAGAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 807
871 GGGAGAGAGCGCTTGAAGAGCGC---CTACCCAGCTCATGAGAGAGGCTCCGC 924
808 GGGGCGCTCAACCAAGAGAGCGCGCTGAGCTGAGCGGAGAGAGAGCTGAGCTGAG 864

```

```

RESULT 15
US-09-726-614-23
Sequence 23, Application US/09726614
Patent No. 6514735
GENERAL INFORMATION:
APPLICANT: Susan E. Jensen
APPLICANT: Kwamena A. Aidoo
APPLICANT: Ashish S. Paraskar
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
Patent No. 6514735
TITLE OF INVENTION: Acid Biosynthesis
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSER: JACOBSON, PRICE, HOLMAN & STERN, PLLC
STREET: The Jennifer Building, 400 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/726,614
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/790,462
FILING DATE: 29-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: D. Douglas Price
REGISTRATION NUMBER: 24,514
REFERENCE/DOCKET NUMBER: 1418/P57452US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-6666
TELEFAX: (202) 39305350
TELEX: RCA 248593 IDEA UR
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1227 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-726-614-23

```

```

Query Match      6.4%; Score 62.2; DB 4; Length 1227;
Best Local Similarity 44.4%; Pred. No. 0.00055;
Matches 345; Conservative 0; Mismatches 423; Indels 9; Gaps 2;

```

```

157 CTGGAGCGCTTCCACCTGAGAGCCCTGCGCGCTTCCCTGAGAGAGTCTTAGG 216
88 CTGGCGGGGCTGCGGTCCAGAGAGCGCGAGCGGGTGAAGCTGTGGGAGCGGAGCGCAG 147

```





**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 13:53:05 ; Search time 66 Seconds  
(without alignments)  
718.508 Million cell updates/sec

Title: US-09-938-901-2

Perfect score: 1700

Sequence: 1 VEAMRKALLAWYRENAAPLP.....VLRKALPLLAHGVPLPDA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 541936

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications-AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1700	100.0	325	11	US-09-938-901-2
2	486.5	28.6	486	9	US-09-925-301-1326
3	377	22.2	313	15	US-10-156-761-13241
4	365	21.5	293	10	US-09-738-626-6433
5	199	11.7	268	10	US-09-864-866-443
6	184.5	10.9	310	15	US-10-156-761-12127
7	147.5	8.7	260	10	US-09-738-626-3828
8	145	8.5	211	9	US-09-912-020-296
9	128	7.5	281	15	US-10-128-714-3128
10	128	7.5	461	15	US-10-128-714-8128
11	112.5	6.6	1332	12	US-09-840-743-11
12	111	6.5	1729	12	US-09-840-743-2
13	110	6.5	1413	12	US-09-840-743-8
14	108.5	6.4	635	12	US-10-302-840A-6
15	104.5	6.1	549	15	US-10-156-761-14029

16	103.5	6.1	298	12	US-09-840-743-33	Sequence 33, App1
17	102	6.0	1133	15	US-10-156-761-11645	Sequence 11645, A
18	100	5.9	807	15	US-10-044-692-5	Sequence 5, App11
19	100	5.9	807	15	US-10-044-539-5	Sequence 5, App11
20	100	5.9	1132	10	US-09-990-080-2	Sequence 2, App11
21	100	5.9	1132	10	US-09-749-728B-31	Sequence 31, App1
22	100	5.9	1132	10	US-09-843-676-225	Sequence 225, App
23	100	5.9	1132	10	US-09-853-052-2	Sequence 2, App1
24	100	5.9	1132	12	US-10-295-681-57	Sequence 57, App1
25	100	5.9	1132	15	US-10-053-758-225	Sequence 225, App
26	100	5.9	1132	15	US-10-208-243-2	Sequence 2, App11
27	100	5.9	1132	15	US-10-054-295-225	Sequence 225, App
28	100	5.9	1132	15	US-10-054-611-225	Sequence 225, App
29	100	5.9	1132	15	US-10-105-963-2	Sequence 2, App11
30	100	5.9	1132	15	US-10-044-692-2	Sequence 2, App11
31	100	5.9	1132	15	US-10-044-539-2	Sequence 323, App
32	100	5.9	1154	15	US-10-044-692-323	Sequence 323, App
33	100	5.9	1154	15	US-10-044-539-323	Sequence 323, App
34	100	5.9	1189	15	US-10-044-692-325	Sequence 325, App
35	100	5.9	1189	15	US-10-044-539-325	Sequence 325, App
36	100	5.9	1200	15	US-10-044-692-324	Sequence 324, App
37	100	5.9	1200	15	US-10-044-539-324	Sequence 324, App
38	100	5.9	1285	15	US-10-044-692-314	Sequence 314, App
39	100	5.9	1285	15	US-10-044-539-314	Sequence 314, App
40	100	5.9	1407	15	US-10-044-692-334	Sequence 334, App
41	100	5.9	1407	15	US-10-044-539-334	Sequence 334, App
42	98	5.8	748	11	US-09-880-505-154	Sequence 154, App
43	98	5.8	748	14	US-10-051-643-154	Sequence 154, App
44	98	5.8	842	9	US-09-815-242-11950	Sequence 11950, A
45	97.5	5.7	412	15	US-10-156-761-8018	Sequence 8018, A

## ALIGNMENTS

RESULT 1  
US-09-938-901-2  
; Sequence 2, Application US/09938901  
; Publication No. US20030008291A1  
; GENERAL INFORMATION:  
; APPLICANT: Yokoyama Shigeyuki  
; TITLE OF INVENTION: GENE ENCODING DNA REPAIR ENZYME  
; FILE REFERENCE: PH-1261-US  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: JP2001-47762  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 325  
; TYPE: PRT  
; ORGANISM: Thermus thermophilus  
US-09-938-901-2

Query Match	100.0%	Score 1700;	DB 11;	Length 325;
Best Local Similarity	100.0%;	Pred. No. 2,6e-152;		
Matches 325;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	VEAMRKALLAWYRENAAPLPWGEKDPYRVLSVEVLLOQTRVQALPYRRFLERPPTLK	60	
DB	1	VEAMRKALLAWYRENAAPLPWGEKDPYRVLSVEVLLOQTRVQALPYRRFLERPPTLK	60	
QY	61	ALAAASLEEVLRWOGAGYVRAEHLRLARSVEELPPSFALRGJLGLGPTAAVAST	120	
DB	61	ALAAASLEEVLRWOGAGYVRAEHLRLARSVEELPPSFALRGJLGLGPTAAVAST	120	
QY	121	AFGEVAAVADGNRVYLSRLFAESPEKELFALAOGLLEPGVDPGVWQALMELGATVC	180	
DB	121	AFGEVAAVADGNRVYLSRLFAESPEKELFALAOGLLEPGVDPGVWQALMELGATVC	180	
QY	181	LPRRPGACPLGAPFCRGKRAEPGRYPAPRRRAKKEETVALVLLGKRGVLELDEGRFQG	240	

```
Db 181 LPKRRPGACGLGAFCCGKEAPGRYPAPRRKRAKEERLVALVLLGRKGHLEERLEGREFGG 240
Qy 241 LYGVPLEPPELPGRREAFVGRSRPLGVEVHALTHRLRVRVRAALWEGEEDPWKRPLP 300
Db 241 LYGVPLEPPELPGRREAFVGRSRPLGVEVHALTHRLRVRVRAALWEGEEDPWKRPLP 300
Qy 301 KLMEKVLKRALPLLAHAGVPLPDA 325
Db 301 KLMEKVLKRALPLLAHAGVPLPDA 325

RESULT 2
US-09-925-301-1326
; Sequence 1326, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/0500/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1326
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (438)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (447)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1326

Query Match 28.6%; Score 486.5; DB 9; Length 486;
Best Local Similarity 36.6%; Pred. No. 1.9e-37;
Matches 136; Conservative 46; Mismatches 107; Indels 83; Gaps 13;

Qy 1 VEAMRKALLAWYRENAARPLPWRG-----EKDPRVLVSEVLLQOTRVEQALFYRRF 52
Db 104 VTAFFGSLISWYDQEKRPDPWRRAEDMDLDRAYAVVSEVMLOQTQVATVINYTYG 163
Qy 53 LERFPTLALAASLEEVLRVWQAGYRRABHLRLARS-VEEL---PPSFAEL-RGL 106
Db 164 MOKWPTLQDLASASLEEVWQMLAGLGYISRGRLDGGAKVVEELGGMHPRTAETLQQL 223
Qy 107 PELGPTYAAVAASIAFGERVAADVGNRRVLSRLFA-----RESPEKEELFALAAGLPE 161
Db 224 PGVGRYTAGAIAISIAFGATGVVDGVAVALCRVAIRAGADPSSTLVSOQLMGLAOQLV-D 282
Qy 162 GVDPGVWNOALMELGATVCLPKRRPGACPLGAFCRGKE----- 200
Db 283 PARPGDFNQAMMELGATVCTPQRPLCSQCPVESLCRAARQVRVBOEDLLASGLSGSPDVEE 342
Qy 201 -P-----GRYPAPRRKRAK---EERLVALV-----LGRKGVHLE 232
Db 343 CAPNTGQCHLCPLPSPPMOTLGAVNFPKASRKRPPRESSSATCYLEOGGALGAQQLLVQ 402
Qy 233 R-LEGRFOGLYGP--LFPPEELPGRFAF-----GVRSPGLGEVNHALTHRL 278
Db 403 RNSGLLAGLWEPSPVTWEPSEQLQKALLQELQRYAGRPATNKHNLGEVYHTSHIKL 462
Qy 279 RVEVGAALMEGE 290
```

```
Db 463 TYGYVGLALEGQ 474

RESULT 3
US-10-156-761-12241
; Sequence 12241, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12241
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12241

Query Match 22.2%; Score 377; DB 15; Length 313;
Best Local Similarity 39.6%; Pred. No. 2.3e-27;
Matches 97; Conservative 32; Mismatches 88; Indels 28; Gaps 7;

Qy 2 EAMRKALLAWYRENAARPLPWR-RGEKDPYRVLVSEVLLQOTRVEQALFYRRFLEPFTLK 60
Db 25 EALHAQVIAWFDENARLDLPWRPRDAGPWGVSEFMLOQTPVNVLLVEYEQMLRWPRPA 84
Qy 61 ALAASLEEVLRVWQAGYRRABHLRLARSVEE-----LPPSFAELRGLPGIGPTTAA 115
Db 85 DLAKRPAGEAVRANGRLGYPRRALRLHGAVALTERHNGDVPTEHAQLLALPGIGEYTA 144
Qy 116 AVASTAFGERVAADVGNRRVLSRLF-----ARSPKEELFALAAGLPGVDPGV 167
Db 145 AVASFAVQORAAVLDFTNVRVFAVAVTGQYTPMATTAAEKK-----LARLLPDESTAS 200
Qy 168 -WNOALMELGATVCLPKRRPGACPLGAFCCGKEA--PGRYPAPRK-----BRAKEER 217
Db 201 RMAAASMLGALVCTAKNETCHRCPIAGCCAMRLAGKPEHNDPPRRGQYAGTOROYGR 260
Qy 218 LVAVL 222
Db 261 LLAVL 265

RESULT 4
US-09-738-626-6433
; Sequence 6433, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASAKO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
```

CURRENT APPLICATION NUMBER: US/09/738, 626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 6433  
LENGTH: 293  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6433

Query Match 21.5%; Score 365; DB 10; Length 293;  
Best Local Similarity 34.6%; Pred. No. 2,9e-26;  
Matches 99; Conservative 47; Mismatches 120; Indels 20; Gaps 7;

OY 3 AMRKALLAWYRENAARPLWGEKDP---RYLVSEVLLOOTRVEQALPYRRFLERPT 58  
DB 5 AFQALLWFRANARDLAMR---DPNTSAMGILLSEVMSOOTPARVEPIWREMERKPT 61  
OY 59 LKALAASLEEVLRWQAGYRRRAEHLRLA-----RSVEELPPFAELRGLGPGPYT 113  
DB 62 PEDPASTDEILRSWCKLIPPRALRLKCAEYIVKHNAGEVDYVALLALPGIDYT 121  
OY 114 AAFAVASTAFGEERAAVGNRRVLSRLPANE---SPREKELFALAGLLPEGVDPGVNQ 170  
DB 122 ARVAARHFGQRPVVDVTVNRVYQRAVAGSYLAGPAKKQELIDVSLPTHTAP-BFSA 180  
OY 171 ALMELGATVCLPRRCGACPLGFCGKFAKGRYPRARRKRAEERLVALVLLGR--KG 228  
DB 181 ALMELGALICTAVSPKCDTCTPLDQCQWKLGCSPSEELASAKKRVQKFTGTDROVRG 240  
OY 229 VHEERLEGRFQ--GLYGVPLFPPEELPGRFAFGVRSRPLGEVHNA 272  
DB 241 LINDVLRNATAPVPLSAIDVWPPDDAQRSLRSLIEDGLAEQNEA 286

RESULT 5  
US-09-864-866-43  
Sequence 43, Application US/09864866  
Patent No. US20020127656A1  
GENERAL INFORMATION:  
APPLICANT: Lloyd, R. Stephen  
APPLICANT: McCullough, Amanda K.  
TITLE OF INVENTION: DNA REPAIR POLYPEPTIDES AND METHODS OF USE  
FILE REFERENCE: 265 00170101  
CURRENT APPLICATION NUMBER: US/09/864, 866  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/206, 279  
PRIOR FILING DATE: 2000-05-23  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 43  
LENGTH: 268  
TYPE: PRT  
ORGANISM: Micrococcus luteus  
US-09-864-866-43

Query Match 11.7%; Score 199; DB 10; Length 268;  
Best Local Similarity 32.4%; Pred. No. 1.2e-10;  
Matches 70; Conservative 31; Mismatches 95; Indels 20; Gaps 8;

OY 27 PYRVLVSEVLLOOT---RVEQALPYRRFLERPTLKLAAASLEEVLRWQAGYRR-R 82  
DB 44 PFELVAATVLSAOTTDVRYNAATP---ALPARPDAHMAAAAEPELQELVIRSTGFYRNK 100  
OY 83 AEHLRLARSV-----ELPPFAELRGLGPGPYTAAAVASIAFGERVAADVGNVRLV 137

DB 101 ASALLRLSQELVGRHDEGVPARLEDVVALPGVGRKTAEVYLGMAFGPGITVDTHFGRLA 160  
OY 138 SRL-FARSPREKELFALAGLLPEGVDPGVWQNALMELGANTVCLPRRCGACPLGFC 196  
DB 161 RRLGPTDTPDGKGR-ARRGRVPPARWTLNLSRLTHGRVRCARRPACGRCPIDRWG 219  
OY 197 ----RGKEAPGRYPARRRRRAK--EERLVALVLLGR 226  
DB 220 PSYAGETDPPERARALLAYELKPGRELELLRAGR 255

RESULT 6  
US-10-156-761-12127  
Sequence 12127, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156, 761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 12127  
LENGTH: 310  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-12127

Query Match 10.9%; Score 184.5; DB 15; Length 310;  
Best Local Similarity 25.8%; Pred. No. 3.5e-09;  
Matches 70; Conservative 44; Mismatches 110; Indels 47; Gaps 10;

OY 13 RENARPLWRCG-----KDPYRVLVSEVLLOOT---RVEQALPYRRFLERPTLKLAAA 65  
DB 65 RELAEVPIYAHPELDFERSPLIATVLSAOTDLRVNQTT---ALFAKPTPEBDLAAA 121  
OY 66 SLEEVLRWQAGYRR-RAEHLRLARSV-----ELPPFAELRGLGPGPYTAAAVAS 119  
DB 122 NPEVEEILRPTGFRKATKSVIGLSKALVEEFGGEVGRLEDLVKLPBGVGRKTAFAVYLG 181  
OY 120 IAFGEERAAVGNRRVLSRLFARESPREKELFALAGLLPEGVDPGVWQNALMEL--G 176  
DB 182 NAFGRPGITVDTHQRLVRRWQMTDEKDPDKIEAVGALFPRK---SEMTLSHHVYFNG 237  
OY 177 ATVCLPRRCGACPLGFCGKFAKGRYPRARRKRAEERLVALVLLGRGVHLERLEG 236  
DB 238 RICHAKRPACGACPIAPLCAYGEGETDPKAKKLKYE-----KG----- 279  
OY 237 RFQGLYGVPLFPPEELPGRFAFGVRSRPLG 267  
DB 280 ---GFPQGRINLPQAYLD---AGGIPAPPLG 304

RESULT 7  
US-09-738-626-3828  
Sequence 3828, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO

APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAOKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738, 626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 3828  
LENGTH: 260  
TYPE: PRT  
ORGANISM: *Corynebacterium glutamicum*  
US-09-738-626-3828

Query Match 8.7%; Score 147.5; DB 10; Length 260;  
Best Local Similarity 29.0%; Pred. No. 8.6e-06;  
Matches 56; Conservative 29; Mismatches 83; Indels 25; Gaps 8;

QY 26 DRYRVLYSEVLLQO---TRVEQALPYRRFLERFPTLKALAASLEEVLRVMOGAGYR- 81  
DB 53 NPLELVATILSAQCTDVRVNGVTP---ALFKRYPTATDVAANDKRELEEFIRPTGFYR 109  
QY 82 RAHLHRLARSV-----ELPPSFAELRGLPGIGPTAAVAASVANGERYAAVDGNVRRV 136  
DB 110 KATSLGLGGLALSLHDGQVPGTLEQLVLELPGVGRKTANVVGNAFGVGITVDTFHGR 169  
QY 137 LSR--LFARSPKEKELFALAGLLEPGVDPGVW---NQALMELGATVCLPKRPGGACP 191  
DB 170 VRRLKTLDEEDPYKVE-----KVMNELIEKPEMTFHSRLIHGRKRCHSRACGACM 223  
QY 192 LGAFGR--GKEAP 202  
DB 224 LAADCPSPGLECP 236

## RESULT 8

US-09-912-020-296  
Sequence 296, Application US/09912020  
Patent No. US20020045592A1  
GENERAL INFORMATION:  
APPLICANT: Zyskind, Judith  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Trawick, John  
APPLICANT: Forsyth, R. Allyn  
APPLICANT: Froelich, Jamie M.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
FILE REFERENCE: ELITRA-001DVI  
CURRENT APPLICATION NUMBER: US/09/912,020  
CURRENT FILING DATE: 2001-07-23  
PRIOR APPLICATION NUMBER: 09/492,709  
PRIOR FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: 60/117,405  
PRIOR FILING DATE: 1999-01-27  
NUMBER OF SEQ ID NOS: 485  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 296  
LENGTH: 211  
TYPE: PRT  
ORGANISM: *E. coli*  
US-09-912-020-296

Query Match 8.5%; Score 145; DB 9; Length 211;  
Best Local Similarity 29.2%; Pred. No. 1.1e-05;  
Matches 64; Conservative 33; Mismatches 88; Indels 34; Gaps 11;

QY 2 EAMRKALWYRE-NRPLPMNGEKDPRVNVSEVLLQO---TRVEQALPYRRFLERFP 57  
DB 3 KAKRLILRLKRNHNHPPTTELNFSSPEFLVALVLSAQATDVSNKAT-----AKLYP 56  
QY 58 ---TKALAASLEEVLRVMOGAGYR-RAHLHRLARSV-----ELPPSFAELRGLPG 108  
DB 57 VANTPAAMELEGEVGYKTYIKTIGLYNSKAENIITCRILLEQHNGEVPEDEAALEALPG 116  
QY 109 LGPYTAAVAASVAFGRVAAVDGNVRVLSRL-FAR---ESPKEKELFALAQGLPE-- 161  
DB 117 VGRKTANVVLNFAFGPTTAVDTHIFRVCNRTQFAPGRKVEQVEEKL-----KVPAPF 171  
QY 162 GVDPGVWQNALMELGATVCLPKRPGGACPLGAFGRKE 200  
DB 172 KVDCHHW---LILHGRYTCIARKPRGSGCIIEDLCYEKE 207

## RESULT 9

US-10-128-714-3128  
Sequence 3128, Application US/10128714  
Publication No. US20030119013A1  
GENERAL INFORMATION:  
APPLICANT: Jiang, Bo  
APPLICANT: Hu, Wenqi  
APPLICANT: Tishkoff, Daniel  
APPLICANT: Zamudio, Carlos  
APPLICANT: Eroschkin, Alexey M  
APPLICANT: Lemieux, Sebastien M  
TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and  
FILE REFERENCE: 10182-018-999  
CURRENT APPLICATION NUMBER: US/10/128,714  
CURRENT FILING DATE: 2002-04-23  
PRIOR APPLICATION NUMBER: US 60/285,697  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: US 60/287,066  
PRIOR FILING DATE: 2001-04-27  
PRIOR APPLICATION NUMBER: US 60/295,890  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: US 60/303,899  
PRIOR FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: US 60/316,362  
PRIOR FILING DATE: 2001-08-31  
NUMBER OF SEQ ID NOS: 8603  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3128  
LENGTH: 281  
TYPE: PRT  
ORGANISM: *Aspergillus fumigatus*  
US-10-128-714-3128

Query Match 7.5%; Score 128; DB 15; Length 281;  
Best Local Similarity 28.6%; Pred. No. 0.00066;  
Matches 44; Conservative 26; Mismatches 66; Indels 18; Gaps 7;

QY 59 LKALAASLEEVLRVMOGAGYRRAHLHRLARSV-----ELPPSFAELRGLPGIGPY 112  
DB 97 LENTLAWSPEKLNELITVGFHNHNTKYIAAAEILIDQYNSDIPSTAEELMKLPGVGPK 156  
QY 113 TAAVAASVAFG-ERVAAVDGNVRVLSRLPARASPKKEKELFALA-QGILLPEGVDPGVW-- 168  
DB 157 MAYICSAANGKDGITVDVHVRIT-TNLMGMHKTPTPESTRNALSLWLPF-----DKWHE 211  
QY 169 -NQALMELGATVCLPKRPGGACPLGA--FCRGR 199  
DB 212 INKLVLGLQGVCLPVGRCGECDDLAGTKLCKSE 245

```

RESULT 10
US-10-128-714-8128
; Sequence 8128, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengf
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshtkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8128
; LENGTH: 461
; TYPE: PRF
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8128

```

```

Query Match
Best Local Similarity 7.5%; Score 128; DB 15; Length 461;
Best Local Similarity 28.6%; Pred. No. 0.0013;
Matches 44; Conservative 26; Mismatches 66; Indels 18; Gaps 7;
QY 59 KALAAASLEEVYRVNQGAGYRRRAHLRLARSV-----EELPPSFARLRPLGPR 112
DB 277 LENTIAVSEPKNELLRITGVFNHNNKTKYIKAAAEILRDYNSDIPSTAEELMKLPVGPK 336
QY 113 TAAVASIAFG-ERVAVDGNVRVLSRLFARESPEKELEFALA-OGLLPEGVDPGVW-- 168
DB 337 MAYLCSSAAMGKDEGCGVHVHNRIT-TNLMGMHKTTPRETTRALSMPLR-----DKMHE 391
QY 169 -NOALMELGATVCLPKRRPGACPLGA--FCRCGK 199
DB 392 INKLIVGLQGTVCLLPYGRRCGECDDLDTGFKLKSE 425

```

```

RESULT 11
US-09-840-743-11
; Sequence 11, Application US/09840743
; Publication No. US20030135890A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Choi, Yeonhee
; APPLICANT: Hannou, Mike
; APPLICANT: Okamuro, Jack Kishiro
; APPLICANT: Tatarinova, Tatiana Valerievna
; TITLE OF INVENTION: The Regents of the University of California
; FILE REFERENCE: 023070-099910US
; CURRENT APPLICATION NUMBER: US/09/840,743
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 09/553,690
; PRIOR FILING DATE: 2000-04-21
; NUMBER OF SEQ ID NOS: 119
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 1332
; TYPE: PRF

```

```

; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: DMT3 (1DMT3)
US-09-840-743-11

```

```

Query Match
Best Local Similarity 6.6%; Score 112.5; DB 12; Length 1332;
Best Local Similarity 22.8%; Pred. No. 0.16;
Matches 59; Conservative 34; Mismatches 95; Indels 71; Gaps 12;
QY 86 LHLARS-----VEELPPSFAR--LRGLPGIGPTAAVASIACGERVAADVRR 135
DB 849 LNLVKKHGSIDLEWLDVDPDKAKEXYLSINGLKSVBCVRLSLHQIAFPVDTVNGR 908
QY 136 VLSRLF--ARESPEKELEFALAOGLLPEGVDPGVW-----NOALMEL-----GAT 178
DB 909 IAVRLGWVLPDLPDELDMHLELYPYESVQXITWPRCLCKDOKITIELHYHMTTGGV 968
QY 179 VCLPKRRPGACPLGAFCR--GKEAPGRYPAP-----RRRAKERLYV---- 219
DB 969 FCTKVRKNCNACPMKACRHYSSARASRLALPEPESDRTSVMIHERSRKRPVYVNR 1028
QY 220 -ALVLLGRKGVHLERLEGRFOGLYGVPLPPEELPGREANFGVRSRPLGEVRLAL-THRR 277
DB 1029 PSFLYOEKQEOEARSO-NCEPIIEEPASPEPEV-----IEHDIEDYPR 1071
QY 278 LRVFVRGALMEGEGEDPWK 296
DB 1072 DKNNV-----GTSEDPWE 1084

```

```

RESULT 12
US-09-840-743-2
; Sequence 2, Application US/09840743
; Publication No. US20030135890A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Choi, Yeonhee
; APPLICANT: Hannou, Mike
; APPLICANT: Okamuro, Jack Kishiro
; APPLICANT: Tatarinova, Tatiana Valerievna
; TITLE OF INVENTION: The Regents of the University of California
; FILE REFERENCE: 023070-099910US
; CURRENT APPLICATION NUMBER: US/09/840,743
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 09/553,690
; PRIOR FILING DATE: 2000-04-21
; NUMBER OF SEQ ID NOS: 119
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1729
; TYPE: PRF
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: DEMETER (DMT)
US-09-840-743-2

```

```

Query Match
Best Local Similarity 6.5%; Score 111; DB 12; Length 1729;
Best Local Similarity 22.3%; Pred. No. 0.31;
Matches 51; Conservative 27; Mismatches 91; Indels 60; Gaps 9;
QY 60 KALAAASLEEVYRVNQGAGY-----YRRRAHLRLARS-----VEELPPSFAR--LR 104
DB 1219 EAIRRASISISIAIRGERNNNNLAVRIKDFLEIVADHGIDLEWLRSPDPKAKOYL 1278
QY 105 GLDGLGPTAAVASIAFGERVAVDGNVRVLSRLF--ARESPEKELEFALAOGLLPE 161
DB 1279 SINGLGKSVCECRRLTLNHLAPVDITNGRIAVRGGWVLPQLPSLDLHLELYPVE 1338
QY 162 GVDPGVW-----NOALMELGATVCLPKRRPGACPLGAFCR--GKEAPGR 204
DB 1339 STOKFLMPRLCKDORTLVELAHYOLITFGKVFCTKSRPNCNACPMRGECRHHFASAVASAR 1398

```

QY 205 YPAERKRRKEERLVALVLLGRKGVHLERLEGRFGVLPPEELP 253  
Db 1399 LALP-----ADERSITSAT-----IPV-PPSEFP 1422

## RESULT 13

US-09-840-743-8  
Sequence 8, Application US/09840743  
Publication No. US20030135890A1  
GENERAL INFORMATION:  
APPLICANT: Fischer, Robert L.  
APPLICANT: Choi, Yeonhee  
APPLICANT: Hannon, Mike  
APPLICANT: Okamoto, Jack Kishiro  
APPLICANT: Tatarkinova, Tatiana Valerievna  
APPLICANT: The Regents of the University of California  
TITLE OF INVENTION: Nucleic Acids That Control Plant Development  
FILE REFERENCE: 023070-09991005  
CURRENT APPLICATION NUMBER: US/09/840,743  
CURRENT FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: US 09/553,690  
PRIOR FILING DATE: 2000-04-21  
NUMBER OF SEQ ID NOS: 119  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8  
LENGTH: 1413  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
OTHER INFORMATION: DMT2 (1DMT2)  
US-09-840-743-8

Query Match 6.5%; Score 110; DB 12; Length 1413;  
Best Local Similarity 23.3%; Pred. No. 0.29;  
Matches 42; Conservative 25; Mismatches 71; Indels 42; Gaps 6;

QY 60 KALAASLEEVLRVMGAGY-VRRREHLRLARSE-----EL 96  
Db 886 KAIRADYKEVAETKSRGMHKLAEKIOYLTLMKIMGFLDLVNDHGSIDLEMLDV 945  
QY 97 PPSFAE-LRGLPGIPYTAAVAASIFGERVAADVGNVRYLSLF--ARESPKEKL 151  
Db 946 PPDKAKEYLLSFNGGLKSVCEVRLTLHLAFLPDVTNGRAVLGVLPQLPESLQL 1005  
QY 153 FALAGGLPEGVDPEGV-----NQALMEL-----GATVCLPKRRGACPLGAFCR 197  
Db 1006 HLEMYPMLESIOXYLMPRLCKLDQKTLIELHYQMITEGKVFCTKSKPNCNACPMKSECR 1065

## RESULT 14

US-10-302-840A-6  
Sequence 6, Application US/10302840A  
Publication No. US20030134794A1  
GENERAL INFORMATION:  
APPLICANT: Madlison, Edwin L.  
APPLICANT: Ong, Edgar O.  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP17, THE ENCODING OF INVENTION: POLYPEPTIDES AND METHODS BASED THEREON  
FILE REFERENCE: 24745-1622  
CURRENT APPLICATION NUMBER: US/10/302,840A  
CURRENT FILING DATE: 2003-01-24  
PRIOR APPLICATION NUMBER: 60/332,015  
PRIOR FILING DATE: 2001-11-20  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6  
LENGTH: 635  
TYPE: PRT  
ORGANISM: Homo Sapien  
FEATURE:  
NAME/KEY: DOMAIN  
LOCATION: (104)...(332)  
OTHER INFORMATION: CVSP17 protease domain

US-10-302-840A-6

Query Match 6.4%; Score 108.5; DB 12; Length 635;  
Best Local Similarity 27.6%; Pred. No. 0.14;  
Matches 66; Conservative 22; Mismatches 108; Indels 43; Gaps 10;

QY 105 GLPGL-SPYTAANAASINIFGERVAADVGNVRYLSRLEFARSPEKELEFALAGLPEGV 163  
Db 314 GEGKPGVYTRYAAVFKMDLOMSAASSRSPSCRELLAMDPQELQADARLCAPYARL 373  
QY 164 DPGVNNQALMELGATVCLPKRRPGACPLGAFCR--GKEAPGRYP-----APRRKRAK 214  
Db 374 CPGS-QGACARLAHQCCQRRRRRC-----GGFCSPRGGRGAEGPGQPLTAPTPVRS 427  
QY 215 EERLVALVLLGRKGVHLERLEGRFGVLPPEELPGREAAFGVRSRPLGFRHALT 274  
Db 428 ELHSLAHTLGLHLR-NAOELGPPRGELR--RLAPALLP-----APALRESPLHARERL 480  
QY 275 HR-----RLREVRGALWMEGEGDPPKRRPLRLMEKYLRLKALPLAHGVPLP 323  
Db 481 HSGRNAQTRFPKRRPFRPGRGANGCPGLEPLRQKLAALQ-----AHAMTILQVP 529

## RESULT 15

US-10-156-761-14029  
Sequence 14029, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 14029  
LENGTH: 549  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-14029

Query Match 6.1%; Score 104.5; DB 15; Length 549;  
Best Local Similarity 21.5%; Pred. No. 0.27;  
Matches 92; Conservative 46; Mismatches 134; Indels 155; Gaps 21;

QY 7 ALLAWYRENAAPLPMWGEKDDPYRVIVSEV-----LLOQT-----RVEQAL----- 46  
Db 112 ALVETARTAGLPLVQHLREVPYTVTEVHTELVNGHVALLDRAEYHNHRCETALLGGG 171  
QY 47 -PYRRFLERF--PLKLAASL-----EVLRVMOGAGYRRRAHLRLARS 92  
Db 172 VPQVLRILADFSGNPFLFETADGRLLYAAGAPADTDPLQWEGC-----RGQH----- 220  
QY 93 VEELPPSFAELRGLPGYTAANAASIAF--GERVAADV-----GNVRYL----- 137  
Db 221 -KDAPPAGTTIVDPGCGAGSVARAVLLLVGNPNVAFVHIAAERAAAGSLAVVLMQAR 279  
QY 138 -----SLFPARESPKEKELFALAOG--LPL-----EGVDP-GVW 168  
Db 280 OEERELAARGDFLTDLAGRIDADAPQAQAVLGFKCGAGPLPLVYVRLADGSLPGGW 339  
QY 169 -----NQALMELGATVCLPKRRPGACPLGAFGRGAEAPGRYAPPKRRAKERLYAL 221  
Db 340 AVLARAIVAELIASIGVPLLVGPPVEGVRVPLLLGLRSR-----ERSAVADVRAAA 390



```

0Y      222 VLLGKGNHLEBGRFGLYGVRLEPPEELRG-----REAAGCVNS 263
           | : : : | : : : | : : : | : : : |
Db      391 LRAGERAGMORPGAO-----PPVVVGVAGGWAASAGLENNAAQTATAAGGLSD 440
           | : : : | : : : | : : : | : : : |
0Y      264 RPLGVRNALNHRRLRVEYRGALWEGEDP-----WKRRPLKMEKVLKATPL 313
           || : : | : : : | : : : | : : : |
Db      441 RPKWDAR-----RLDIDLMLRLR--DDPDLAAFVDRDIAIGRLRDHNRKRPILLPTLO 491
           | : : | : : : | : : : | : : : |
0Y      314 --LAHAG 318
           |||||
Db      492 TYLAHAG 498

```

Search completed: September 15, 2003, 14:02:48  
Job time : 69 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 13:40:14 ; Search time 83 Seconds  
(without alignments)  
621.520 Million cell updates/sec

Title: US-09-938-901-2  
Perfect score: 1700  
Sequence: 1 VEAMRKALIAWYRENARPLP.....VLRKALPLAHGVLPDPA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1700	100.0	325	24	ABP56413
2	492	28.9	387	23	ABBS4155
3	491	28.9	349	24	ABP79054
4	486.5	28.6	486	21	ABAB3881
5	484.5	28.5	535	18	AAW31912
6	484.5	28.5	535	19	AAW60243
7	478	28.1	365	23	ABBA9175
8	477	28.1	375	20	AAV35003
9	471.5	27.7	350	22	AAV69755

10	465.5	27.4	391	20	AAV09118
11	465	27.4	384	23	ABP27945
12	464.5	27.3	350	19	AAW60244
13	461	27.1	374	23	ABP30125
14	460.5	27.1	341	20	AAV37104
15	459.5	27.0	391	24	ABU01657
16	443	26.1	360	23	ABP39839
17	442.5	26.0	328	23	ABP65978
18	442	26.0	319	20	AAV09119
19	442	26.0	447	19	AAV86014
20	437	25.7	322	18	AAW27942
21	437	25.7	322	18	AAW25524
22	368	21.6	333	23	ABP27944
23	365	21.5	293	22	AAW26679
24	332.5	19.6	292	22	AAU41327
25	321.5	18.9	292	21	AAW41331
26	321.5	18.9	292	23	ABP7225
27	274.5	16.1	221	18	AAW11703
28	274.5	16.1	221	19	AAW81465
29	274.5	16.1	221	19	AAW48738
30	213.5	12.6	220	23	ABU51793
31	200	11.8	223	22	AAW86207
32	199	11.7	268	23	AAE15907
33	183.5	10.8	218	23	ABP25507
34	178.5	10.5	219	23	ABBA8015
35	176.5	10.4	90	23	ABP35548
36	172.5	10.1	228	23	ABP65491
37	164.5	9.7	198	23	ABU51400
38	164.5	9.7	218	23	ABBS4387
39	164.5	9.7	304	18	AAW23155
40	164	9.6	210	22	AAW81939
41	164	9.6	224	22	AAW82795
42	162	9.5	210	23	ABP25506
43	161.5	9.5	143	22	AAU29544
44	160	9.4	224	23	ABP38380
45	159.5	9.4	367	21	AAW17150

## ALIGNMENTS

RESULT 1	
ABP56413	
ID	ABP56413 standard; Protein: 325 AA.
XX	
AC	ABP56413:
XX	
DT	12-MAR-2003 (first entry)
XX	
DE	Thermus thermophilus DNA repair enzyme MutY protein SEQ ID NO:2.
XX	
KW	Thermus thermophilus; DNA repair enzyme; enzyme; MutY; Recf; Recf;
KW	TRCF; biochemistry; molecular biology; research.
XX	
OS	Thermus thermophilus.
XX	
PN	JF2002247985-A.
XX	
PD	03-SEP-2002.
XX	
PF	23-FEB-2001; 2001JP-0047762.
XX	
PR	23-FEB-2001; 2001JP-0047762.
XX	
PA	(RIKA) RIKAGAKU KENKYUSHO.
XX	
DR	WPI; 2003-078924/08.
XX	
XX	N-PDB; ABZ22143.
XX	
PT	A DNA repair enzyme gene, a protein, a recombinant vector, a
PT	transformant, preparation of DNA repair enzyme, repairing the error
PT	sequence of a DNA, and prevention of error synthesis of a DNA sequence
PT	

XX Claim 1; Page 14-15; 41pp; Japanese.  
PS  
XX AB22243 to AB22146 encode the *Thermus thermophilus* DNA repair enzymes  
CC Muty, Recf, Recf, and TRCF and TRCF from ABP56413 to ABP5416. The enzymes  
CC can be used as research reagents for biochemistry and molecular biology.  
XX  
SQ Sequence 325 AA;

Query Match 100.0%; Score 1700; DB 24; Length 325;  
Best Local Similarity 100.0%; Pred. No. 1.5e-153;  
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEAMRKALLAMYRENAAPLPWGEKDPYRVLYSEVLLQOTRVQALPYRRFLERPTLK 60  
DB 1 VEAMRKALLAMYRENAAPLPWGEKDPYRVLYSEVLLQOTRVQALPYRRFLERPTLK 60  
QY 61 ALAASLEEVLRWOGAGYRRAEHLRLARSVEELPPSFAELRGIPGIPYAAAVAST 120  
DB 61 ALAASLEEVLRWOGAGYRRAEHLRLARSVEELPPSFAELRGIPGIPYAAAVAST 120  
QY 121 AFGEERVAADVGNVRRVLSRLFARESPEKELFALAAGLLPEGVDPGVMMQALMELGATVC 180  
DB 121 AFGEERVAADVGNVRRVLSRLFARESPEKELFALAAGLLPEGVDPGVMMQALMELGATVC 180  
QY 181 LPRRPGACPLGAFGRGKEAPGRYPAPRRRAKEERLVALVLLGRKGVHLERLEGFG 240  
DB 181 LPRRPGACPLGAFGRGKEAPGRYPAPRRRAKEERLVALVLLGRKGVHLERLEGFG 240  
QY 241 LVGVPLPEPELPGRAAAGVSRPLGEVRRHALTHRLRVEYRGALMBEGEDPMKRP 300  
DB 241 LVGVPLPEPELPGRAAAGVSRPLGEVRRHALTHRLRVEYRGALMBEGEDPMKRP 300  
QY 301 KLMEXVLRKALPLLAHAGVPLPDA 325  
DB 301 KLMEXVLRKALPLLAHAGVPLPDA 325

## RESULT 2

ABBS4155  
ID ABBS4155 standard; Protein; 387 AA.

XX  
AC ABBS4155;

DT 16-MAY-2002 (first entry)

XX Lactococcus lactis protein muty.

DE Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

XX Lactococcus lactis IL1403.

OS  
XX  
PN FR2807446-A1.

XX  
PD 12-OCT-2001.

XX  
PF 11-APR-2000; 2000FR-0004630.

XX  
PR 11-APR-2000; 2000FR-0004630.

XX  
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.

XX  
PI Bojoline A, Sorokline A, Renault P, Ehrlich SD;

XX  
DR WPI; 2002-043418/06.

XX  
PT New nucleotide sequence useful in the identification or Lactococcus

XX  
PS Lactis and related species -

XX  
CC Claim 6; SEQ ID No 857; 2504pp; French.

XX  
CC The present invention is related to a Lactococcus lactis nucleotide

CC  
CC sequence (AB90521) and related proteins (ABBS3300-ABBS5621). The

CC nucleic acid sequence is useful in the detection and/or amplification of  
CC nucleic acid sequence, particularly to identify Lactococcus lactis or  
CC related species. The proteins of the invention are useful for the  
CC biosynthesis or biodegradation of a composition of interest. The  
CC invention helps research in lactic bacteria, particularly useful in the  
CC production of yogurt and cheese.  
CC Note: The sequence data for this patent is based on equivalent patent  
CC WO200177334 (published 18-OCT-2001) which is available in electronic  
CC format directly from WIPO at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
XX  
SQ Sequence 387 AA;

Query Match 28.9%; Score 492; DB 23; Length 387;  
Best Local Similarity 39.3%; Pred. No. 3.1e-38;  
Matches 108; Conservative 57; Mismatches 84; Indels 26; Gaps 10;

QY 1 VEAMRKALLAMYRENAAPLPWGEKDPYRVLYSEVLLQOTRVQALPYRRFLERPTLK 60  
DB 8 IKEFGQDLSMTDDOKKPLPKRKTEPEYKISTMSQOYEVMPYERPMKPYTIE 67  
QY 61 ALAASLEEVLRWOGAGYRRAEHLRLARSVEE---LPSFAELRGIPGIPYAA 115  
DB 68 TLAQADDAELLKMEGLGYSRARNLKIAQEVVDKYNKFPDNLADILSLKIGPYTA 127  
QY 116 AVASTAFGERVAADVGNVRRVLSRLFARESPEKELFALAAGLLPEGVD---PGV 168  
DB 128 ALASTSFCALPAIDGNLMRYTSRLFEEDCDISKSSSKTF---DGLRKLISKRRGDF 184  
QY 169 NOALMELGATVCLPKRPGACPLGAF---RCKEAPGRYPAPRR-RAKEERLVALV 223  
DB 185 NOALMDLGLSTVCSKSPKCEACPLNYCAAAAGSKL---NYPKTKIKKQKDLTYAFAL 242  
QY 224 LGRKG-VHLERLEGFGIYG-VLPPEPELPGR 256  
DB 243 ENSIGEYTLERPSK--GLADMTFPLTELPAAD 275

## RESULT 3

ABP79054  
ID ABP79054 standard; Protein; 349 AA.

XX  
AC ABP79054;

DT 07-MAR-2003 (first entry).

XX N. gonorrhoeae amino acid sequence SEQ ID 4638.

DE Antibacterial; infection; vaccine; gene therapy.

XX  
OS  
XX  
PN Netisseria gonorrhoeae.

XX  
PD WO200279243-A2.

XX  
PF 10-OCT-2002.

XX  
PR 12-FEB-2002; 2002WO-1B02069.

XX  
PR 12-FEB-2001; 2001GB-0003424.

XX  
PA (CHIR-) CHIRON SPA.

XX  
PI Fontana MR, Pizsa M, Maignani V, Monaci E;

XX  
DR WPI; 2003-058415/05.

XX  
DR N-PSDB; AB240024.

XX  
PT New protein from Netisseria gonorrhoeae, useful for the manufacture of a

XX  
PS medicament for treating or preventing N. gonorrhoeae infection -

XX  
PS Disclosure; Page 528; 815pp; English.

XX  
CC The present invention relates to proteins from Netisseria gonorrhoeae.

XX  
CC Also disclosed are the nucleic acid molecules encoding the proteins and

CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
 CC molecules of the invention.

XX Sequence 349 AA;

Query Match 28.9%; Score 491; DB 24; Length 349;  
 Best Local Similarity 38.1%; Pred. No. 3,4e-38;  
 Matches 128; Conservative 56; Mismatches 104; Indels 48; Gaps 15;

QY 8 LLAAYENAR-PLPMGEKDPYRVLYSEVLLQOTRVQALPYRRLEPPTLKALAAS 66  
 Db LIMQKQNGHNLPMQ-VKNPYCVWLSEIMLQOTQVAADVLYPRFLEKPTVQTAAAP 72  
 QY 67 LEEVLNVOGAGYYRRAEHLRLARSV---EELPPFAELRGLPGIGPYTAAVASIA 121  
 Db 73 ODEVLSLMAGLGYGRRARLNHKAQOIVQGFSTPSEKRDLETLGCVGRSTAAISAF 132  
 QY 122 FGERVAADVGNVRRVLSRLFARE-SPEKE---LRLAAGLLP-EGVDPGVNQALMEL 175  
 Db 133 FNRRTLDGNVRRVLCRVFAODGNPDCKFENSMTLAEISLMPSENAIDPPTYOGIMDL 192  
 QY 176 GATVCLPRKPRGACPLGAFRCGRKEAGRYPARKRRAKEER---LVATLVLRKG-VHL 231  
 Db 193 GATVCKRTKPLCQCMADCEAKKONRTAELPRKTALEVOQLPLVLYVRNDGAILL 252  
 QY 232 ER--LEGROGLGVPLPPEELPGRE--AAGVSRPLGE---VNAHTHRLRLRYEVR 283  
 Db 253 EKRTAKIGMGLCYVCF--ESLNGLSDFPAKLSLTMADMDDEQTALTRHRLHMLT-- 308  
 QY 284 GALMEGEGEDP-----WKRP-----LPLKME 304  
 Db 309 ---PREGQMPSEHNSDGIWKIPGHLKDYGLPKPLE 340

RESULT 4  
 AAB43881  
 ID AAB43881 standard; Protein: 486 AA.

AC AAB43881;  
 DT 08-FEB-2001 (first entry)

DE Human cancer associated protein sequence SEQ ID NO:1326.

XX Human: cancer associated gene; cancer antigen; detection; cancer;  
 KW diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;  
 KW antiinflammatory; antitumor; antineoplastic; antitubercular; antiviral;  
 KW antiinflammatory; antitubercular; antiallergic; antibacterial; cardiant;  
 KW dermatological; neuroprotective; thrombolytic; coagulant; neurotropic;  
 KW vasotropic; antiproliferative; angiogenic; gene therapy; inflammation;  
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
 KW allergic reaction; graft versus host disease; organ rejection;  
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
 KW neurological disease; drug screening.

OS Homo sapiens.

PN MO20055350-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000MO-US05882.

PR 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX

DR WPI: 2000-587533/55.

DR N-PSDB: AAC78090.

PT Novel isolated nucleic acids comprising sequences encoding peptides

XX useful for treating or diagnosing e.g. cancer -

XX Claim 11; Page 1976-1978; 2352pp; English.

CC AAC77607 to AAC78448 encode the human cancer associated proteins given  
 CC in AAB43398 to AAB44238. The proteins can have activities based on the  
 CC tissues and cells the genes are expressed in. Example of activities  
 CC include: cytostatic; proliferative; vulnerability; immunomodulator;  
 CC antidiabetic; antitubercular; antineoplastic; antitubercular;  
 CC antiinflammatory; antitubercular; antiallergic; antibacterial; antiviral;  
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
 CC neurotropic; vasotropic; antiproliferative; angiogenic; gene therapy;  
 CC polynucleotides and polypeptides can be used for preventing, treating or  
 CC ameliorating medical conditions and diagnosing pathological conditions.  
 CC Polynucleotides, antibodies, agonists and antagonists from  
 CC the present invention may be used to treat immune disorders by activating  
 CC or inhibiting the proliferation, differentiation or mobilisation of  
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
 CC disorders, allergic reactions, graft versus host disease and organ  
 CC rejection, modulate haemostatic or thrombolytic activity, modulate  
 CC inflammation, cancers, cardiovascular disorders, neurological disease and  
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
 CC the present invention.

XX Sequence 486 AA;

Query Match 28.6%; Score 486.5; DB 21; Length 486;  
 Best Local Similarity 36.6%; Pred. No. 1.4e-37;  
 Matches 136; Conservative 46; Mismatches 107; Indels 83; Gaps 13;

QY 1 VEAMRKALLMAYRENAARLPMPRG-----EKDPYRVLYSEVLLQOTRVQALPYRRF 52  
 Db 104 VTAFRGSLSMYDOEKDRLRRRAEDMDLDRAYVWVSEVLLQOTQVAATVNTYTTGW 163  
 QY 53 LEEFPTLKALAASLEEVNQLMAGLGYYSRGRRLQEGARVYBELGSHMPRTAETLQQL 223  
 Db 164 MOKWPTLQDLASASLEEVNQLMAGLGYYSRGRRLQEGARVYBELGSHMPRTAETLQQL 223  
 QY 107 PGLGPTTAAAVASTAEBERAAVNGNRRVLSRLFA-----RESPKKEKFLAAGLLPE 161  
 Db 224 PGVGRYTAGAIASTAFQATGVVDGNVARYLCRVRAAGADPSSTLYSQDLMLAQQLV-D 282  
 QY 162 GVDPGVWVNOALMELGATVCLPKPRPGACPLGAFRCRKE----- 200  
 Db 283 PARPGDFNOAMELGATVCTPQRLCQCVESICRAKQRYEDQLASSLSGSPVEE 342  
 QY 201 -AP-----GRYPARRRRRAK---EERLVALLV-----LGRKGVHLE 232  
 Db 343 CAPRTGCGHCLPSPSEMDQTLGVNFPKRASRKPRESSATCVLEQPALGQQLLVQ 402  
 QY 233 R-LEGROGLGVPLPPEELPGREAF-----GVNSRPLGEVRRALTRRL 278  
 Db 403 RPNSSGLLAGLMEPSPVTPERSEQLORRALLQELQXAGPLPATXHRILGEVHTFSIKL 462  
 QY 279 RVEVRGALMEGE 290  
 Db 463 TYQVYGLALEGQ 474

RESULT 5  
 AAM31912  
 ID AAM31912 standard; Protein: 535 AA.

AC AAM31912;

DT 27-MAR-1998 (first entry)

XX

DE Human mismatch repair protein Muty.  
XX  
XX Muty; hMYH gene; mismatch repair; non-polyposis colon cancer;  
KW xeroderma pigmentosum; gene therapy; diagnosis; human.  
XX  
OS Homo sapiens.  
XX  
XX MO9733903-A1.  
XX  
XX 18-SEP-1997.  
XX  
XX 11-MAR-1996; 96WO-US03239.  
XX  
XX 11-MAR-1996; 96WO-US03239.  
XX  
XX 11-MAR-1996; 96WO-US03239.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX (SMIK ) SMITHKLINE BEECHAM CORP.  
XX  
XX Wet Y:  
XX  
XX WPI: 1997-470811/43.  
XX DR N-PSDB; AAT89194.  
XX  
XX  
XX DNA encoding human Muty protein - useful for detecting and treating  
PT mismatches in DNA especially in non-polyposis colon cancer and  
PT xeroderma pigmentosum  
XX  
XX  
XX Claim 1; Page 39-40; 59pp; English.  
XX  
XX  
XX This protein comprises human Muty, which is homologous to the  
CC Escherichia coli Muty protein involved in the pathway that corrects  
CC A/G and A/C mismatches as well as adenines paired with  
CC 7,8-dihydro-8-oxo-deoxyuracine in mutated DNA. Its amino acid  
CC sequence was deduced from a human cerebellum cDNA clone (see  
CC AAT89194). Muty polypeptides can be expressed in transformed  
CC host cells. The polypeptides, or the polynucleotides encoding  
CC them, can be used e.g. to repair oxidative damage to DNA, to  
CC prevent mutations from oxidative lesions, to treat genetic diseases  
CC related to a mutated hMYH gene, e.g. xeroderma pigmentosum and  
CC neoplasia, and to diagnose an abnormal transformation or a  
CC susceptibility to abnormal transformation of cells, particularly  
CC in a non-polyposis colon cancer.  
XX  
XX  
SQ Sequence 535 AA:  
  
Query Match 28.5%; Score 484.5; DB 18; Length 535;  
Best Local Similarity 36.6%; Pred. No. 2.5e-37;  
Matches 136; Conservative 46; Mismatches 107; Indels 83; Gaps 13;  
  
QY 1 VEAMRKALLAWYRENAAPLPWRG-----EKDPYRVLYSEVLLQOTRVBQALPYRRF 52  
DB 79 VTAIFRGSILSWYDQEKRDLPWRRAEDMDLDRAYAVWSEVMIQOTVATVINYTGW 138  
QY 53 LERFPTLKALAAALEEVLVWOGAGYVRAEHLRLARS-VEEL---PPSFABL--RGL 106  
DB 139 MOKWPTLDDLASASLEEVNQMAGLGYISRGRRLOEGARKVVEELGHNMPRTAETLQQL 198  
QY 107 PGLGPTAAVAASIAFGEVAVADGNVRVLSRLFA-----RESPKEKEFLAQLGLPE 161  
DB 199 PEGVGTAGALASIAFGATGVVDGNVAVLCRVRAIGADPSSSTLVSOQLMGLAOQLV-D 257  
QY 162 GVDPGVWNOALMELGATVCLPKRRPGGACPLGACRGE----- 200  
DB 258 PARPGFNOAMELGATVCTPQRPILCSQCPVESLCAARQRYBQEQLLASGSLSGSPVDEE 317  
QY 201 -APG-----GRYPAPKRRRAK---ERLVALYL-----IGRKGVHLE 232  
DB 318 CAPNTGQCHLCLRPSEPMDOITLGVVNFPRKASRKPRRESSATVCVLEQPGALGAQILLVQ 377  
QY 233 R-LEGRFOGLYGV--LFPPEELPGRBAF-----GVSRPPLGVRNHALTHRRRL 278  
DB 378 RPNSGLLAGLWEPSPVWEPSEBQLOKRALLOELQRAWAGPLPATHLRHLEGVVHTFSHIKL 437

QY 279 RVEVRGALWEGE 290  
DB 438 TYQVYGLALEGQ 449  
  
RESULT 6  
ID AAM60243 standard; Protein; 535 AA.  
XX AAM60243:  
XX  
XX 19-AUG-1998 (first entry)  
XX  
XX Amino acid sequence of human MYH (hMYH).  
DE  
XX Human; MYH; diagnosis; cancer.  
XX  
XX Homo sapiens.  
XX  
XX  
XX Location/Qualifiers  
FH Key 114  
FT Misc-difference 114 /note= "nucleotides encoding this residue not given"  
XX  
XX JP10057076-A.  
XX  
XX 03-MAR-1998.  
XX  
XX 11-MAR-1997; 97JP-0099540.  
XX  
XX 11-MAR-1996; 96US-0013132.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX WPI: 1998-210407/19.  
XX DR N-PSDB; AAV35701.  
XX  
XX New isolated polypeptide - useful for, e.g. diagnosis of cancer  
PT  
XX  
XX Claim 1; Pages 20-21; 23pp; Japanese.  
XX  
XX The present sequence represents human MYH (hMYH). The MYH polypeptide  
CC can be used for diagnosing cancer, where the decreasing levels of the  
CC polypeptide are measured in a sample from a patient.  
XX  
XX  
SQ Sequence 535 AA:  
  
Query Match 28.5%; Score 484.5; DB 19; Length 535;  
Best Local Similarity 34.9%; Pred. No. 2.5e-37;  
Matches 130; Conservative 49; Mismatches 110; Indels 83; Gaps 11;  
  
QY 1 VEAMRKALLAWYRENAAPLPWRG-----EKDPYRVLYSEVLLQOTRVBQALPYRRF 52  
DB 79 VTAIFRGSILSWYDQEKRDLPWRRAEDMDLDRAYAVWSEVMIQOTVATVINYTGW 138  
QY 53 LERFPTLKALAAALEEVLVWOGAGYVRAEHLRLARS-VEEL---PPSFABL--RGL 106  
DB 139 MOKWPTLDDLASASLEEVNQMAGLGYISRGRRLOEGARKVVEELGHNMPRTAETLQQL 198  
QY 107 PGLGPTAAVAASIAFGEVAVADGNVRVLSRLFA-----RESPKEKEFLAQLGLPE 161  
DB 199 PEGVGTAGALASIAFGATGVVDGNVAVLCRVRAIGADPSSSTLVSOQLMGLAOQLV-D 257  
QY 162 GVDPGVWNOALMELGATVCLPKRRPGGACPLGACRGE----- 200  
DB 258 PARPGFNOAMELGATVCTPQRPILCSQCPVESLCAARQRYBQEQLLASGSLSGSPVDEE 317  
QY 201 -APG-----GRYPAPKRRRAK---ERLVALYL-----IGRKGVHLE 234  
DB 318 CAPNTGQCHLCLRPSEPMDOITLGVVNFPRKASRKPRROSSATVCVLEQPGALGAQILLVQ 377  
QY 235 ---EGRFOGLYGV--LFPPEELPGRBAF-----GVSRPPLGVRNHALTHRRRL 278  
DB 378 RPNSGLLAGLWEPSPVWEPSEBQLOKRALLOELQRAWAGPLPATHLRHLEGVVHTFSHIKL 437



malchies 11/; Conservative 44; Mismatches 112

Indels 30; Gaps 7;

22 EALKKWEKKNRSLPWRDNPPTPYSVWSEVMLQOTRAEVVIDYFNOMMERPTTETIAA 81

82 KEEDVIKLEWGLGYSSRAHILTEGARMVMEFEHFKTPDATSI NOIRVCNVTUATAT 14

[illegible][illegible]

7/ 2/ 3/ 4/ 5/ 6/ 7/ 8/ 9/ 10/ 11/ 12/ 13/ 14/ 15/ 16/ 17/ 18/ 19/ 20/ 21/ 22/ 23/ 24/ 25/ 26/ 27/ 28/ 29/ 30/ 31/ 32/ 33/ 34/ 35/ 36/ 37/ 38/ 39/ 40/ 41/ 42/ 43/ 44/ 45/ 46/ 47/ 48/ 49/ 50/ 51/ 52/ 53/ 54/ 55/ 56/ 57/ 58/ 59/ 60/ 61/ 62/ 63/ 64/ 65/ 66/ 67/ 68/ 69/ 70/ 71/ 72/ 73/ 74/ 75/ 76/ 77/ 78/ 79/ 80/ 81/ 82/ 83/ 84/ 85/ 86/ 87/ 88/ 89/ 90/ 91/ 92/ 93/ 94/ 95/ 96/ 97/ 98/ 99/ 100/ 101/ 102/ 103/ 104/ 105/ 106/ 107/ 108/ 109/ 110/ 111/ 112/ 113/ 114/ 115/ 116/ 117/ 118/ 119/ 120/ 121/ 122/ 123/ 124/ 125/ 126/ 127/ 128/ 129/ 130/ 131/ 132/ 133/ 134/ 135/ 136/ 137/ 138/ 139/ 140/ 141/ 142/ 143/ 144/ 145/ 146/ 147/ 148/ 149/ 150/ 151/ 152/ 153/ 154/ 155/ 156/ 157/ 158/ 159/ 160/ 161/ 162/ 163/ 164/ 165/ 166/ 167/ 168/ 169/ 170/ 171/ 172/ 173/ 174/ 175/ 176/ 177/ 178/ 179/ 180/ 181/ 182/ 183/ 184/ 185/ 186/ 187/ 188/ 189/ 190/ 191/ 192/ 193/ 194/ 195/ 196/ 197/ 198/ 199/ 200/ 201/ 202/ 203/ 204/ 205/ 206/ 207/ 208/ 209/ 210/ 211/ 212/ 213/ 214/ 215/ 216/ 217/ 218/ 219/ 220/ 221/ 222/ 223/ 224/ 225/ 226/ 227/ 228/ 229/ 230/ 231/ 232/ 233/ 234/ 235/ 236/ 237/ 238/ 239/ 240/ 241/ 242/ 243/ 244/ 245/ 246/ 247/ 248/ 249/ 250/ 251/ 252/ 253/ 254/ 255/ 256/ 257/ 258/ 259/ 260/ 261/ 262/ 263/ 264/ 265/ 266/ 267/ 268/ 269/ 270/ 271/ 272/ 273/ 274/ 275/ 276/ 277/ 278/ 279/ 280/ 281/ 282/ 283/ 284/ 285/ 286/ 287/ 288/ 289/ 290/ 291/ 292/ 293/ 294/ 295/ 296/ 297/ 298/ 299/ 300/ 301/ 302/ 303/ 304/ 305/ 306/ 307/ 308/ 309/ 310/ 311/ 312/ 313/ 314/ 315/ 316/ 317/ 318/ 319/ 320/ 321/ 322/ 323/ 324/ 325/ 326/ 327/ 328/ 329/ 330/ 331/ 332/ 333/ 334/ 335/ 336/ 337/ 338/ 339/ 340/ 341/ 342/ 343/ 344/ 345/ 346/ 347/ 348/ 349/ 350/ 351/ 352/ 353/ 354/ 355/ 356/ 357/ 358/ 359/ 360/ 361/ 362/ 363/ 364/ 365/ 366/ 367/ 368/ 369/ 370/ 371/ 372/ 373/ 374/ 375/ 376/ 377/ 378/ 379/ 380/ 381/ 382/ 383/ 384/ 385/ 386/ 387/ 388/ 389/ 390/ 391/ 392/ 393/ 394/ 395/ 396/ 397/ 398/ 399/ 400/ 401/ 402/ 403/ 404/ 405/ 406/ 407/ 408/ 409/ 410/ 411/ 412/ 413/ 414/ 415/ 416/ 417/ 418/ 419/ 420/ 421/ 422/ 423/ 424/ 425/ 426/ 427/ 428/ 429/ 430/ 431/ 432/ 433/ 434/ 435/ 436/ 437/ 438/ 439/ 440/ 441/ 442/ 443/ 444/ 445/ 446/ 447/ 448/ 449/ 450/ 451/ 452/ 453/ 454/ 455/ 456/ 457/ 458/ 459/ 460/ 461/ 462/ 463/ 464/ 465/ 466/ 467/ 468/ 469/ 470/ 471/ 472/ 473/ 474/ 475/ 476/ 477/ 478/ 479/ 480/ 481/ 482/ 483/ 484/ 485/ 486/ 487/ 488/ 489/ 490/ 491/ 492/ 493/ 494/ 495/ 496/ 497/ 498/ 499/ 500/ 501/ 502/ 503/ 504/ 505/ 506/ 507/ 508/ 509/ 510/ 511/ 512/ 513/ 514/ 515/ 516/ 517/ 518/ 519/ 520/ 521/ 522/ 523/ 524/ 525/ 526/ 527/ 528/ 529/ 530/ 531/ 532/ 533/ 534/ 535/ 536/ 537/ 538/ 539/ 540/ 541/ 542/ 543/ 544/ 545/ 546/ 547/ 548/ 549/ 550/ 551/ 552/ 553/ 554/ 555/ 556/ 557/ 558/ 559/ 560/ 561/ 562/ 563/ 564/ 565/ 566/ 567/ 568/ 569/ 570/ 571/ 572/ 573/ 574/ 575/ 576/ 577/ 578/ 579/ 580/ 581/ 582/ 583/ 584/ 585/ 586/ 587/ 588/ 589/ 590/ 591/ 592/ 593/ 594/ 595/ 596/ 597/ 598/ 599/ 600/ 601/ 602/ 603/ 604/ 605/ 606/ 607/ 608/ 609/ 610/ 611/ 612/ 613/ 614/ 615/ 616/ 617/ 618/ 619/ 620/ 621/ 622/ 623/ 624/ 625/ 626/ 627/ 628/ 629/ 630/ 631/ 632/ 633/ 634/ 635/ 636/ 637/ 638/ 639/ 640/ 641/ 642/ 643/ 644/ 645/ 646/ 647/ 648/ 649/ 650/ 651/ 652/ 653/ 654/ 655/ 656/ 657/ 658/ 659/ 660/ 661/ 662/ 663/ 664/ 665/ 666/ 667/ 668/ 669/ 670/ 671/ 672/ 673/ 674/ 675/ 676/ 677/ 678/ 679/ 680/ 681/ 682/ 683/ 684/ 685/ 686/ 687/ 688/ 689/ 690/ 691/ 692/ 693/ 694/ 695/ 696/ 697/ 698/ 699/ 700/ 701/ 702/ 703/ 704/ 705/ 706/ 707/ 708/ 709/ 710/ 711/ 712/ 713/ 714/ 715/ 716/ 717/ 718/ 719/ 720/ 721/ 722/ 723/ 724/ 725/ 726/ 727/ 728/ 729/ 730/ 731/ 732/ 733/ 734/ 735/ 736/ 737/ 738/ 739/ 740/ 741/ 742/ 743/ 744/ 745/ 746/ 747/ 748/ 749/ 750/ 751/ 752/ 753/ 754/ 755/ 756/ 757/ 758/ 759/ 760/ 761/ 762/ 763/ 764/ 765/ 766/ 767/ 768/ 769/ 770/ 771/ 772/ 773/ 774/ 775/ 776/ 777/ 778/ 779/ 780/ 781/ 782/ 783/ 784/ 785/ 786/ 787/ 788/ 789/ 790/ 791/ 792/ 793/ 794/ 795/ 796/ 797/ 798/ 799/ 800/ 801/ 802/ 803/ 804/ 805/ 806/ 807/ 808/ 809/ 810/ 811/ 812/ 813/ 814/ 815/ 816/ 817/ 818/ 819/ 820/ 821/ 822/ 823/ 824/ 825/ 826/ 827/ 828/ 829/ 830/ 831/ 832/ 833/ 834/ 835/ 836/ 837/ 838/ 839/ 840/ 84

## APPENDIX 6

●  
●  
●  
●  
●  
●  
●  
●

DNA damage: human xeroderma pigmentosum; complementation; DNA damage: binding protein; nuclease; DNA mismatch; cancer;

n/s-specific adenine glycosylase; synthetic T4 endonuclease V; T4 endo V  
thymine DNA-glycosylase; T4 endo V

*Escherichia coli*.

04-OCT-2001.

28-MAR-2000: 2000MS-102764P

( REGC ) UNIV CALIFORNIA.

WPI; 2001-656920/75.

mutations, e.g. in disease diagnosis, for detecting and quantifying DNA

disclosure; page 98-99; 128pp; English.

sequence 330 AA,

Matches 121; Conservative 61; Mismatches 120; Indels 43;

Db 10 VLDWYDKYGRKTLPMQIDKTPYKVLSEVMLLOOTOVATVIPYFEREMAREPTVTNDI.ANAP 69

70 IDEVT.HITWTGYVABRNHKAACOVNTIHCCKEDEMEEVAATDVCCECTCATCAG

120 CURRENT DOMESTIC ACCOUNTING TREATMENT

1.  $\frac{1}{2}$  2.  $\frac{1}{2}$  3.  $\frac{1}{2}$  4.  $\frac{1}{2}$  5.  $\frac{1}{2}$  6.  $\frac{1}{2}$  7.  $\frac{1}{2}$  8.  $\frac{1}{2}$  9.  $\frac{1}{2}$  10.  $\frac{1}{2}$  11.  $\frac{1}{2}$  12.  $\frac{1}{2}$  13.  $\frac{1}{2}$  14.  $\frac{1}{2}$  15.  $\frac{1}{2}$  16.  $\frac{1}{2}$  17.  $\frac{1}{2}$  18.  $\frac{1}{2}$  19.  $\frac{1}{2}$  20.  $\frac{1}{2}$  21.  $\frac{1}{2}$  22.  $\frac{1}{2}$  23.  $\frac{1}{2}$  24.  $\frac{1}{2}$  25.  $\frac{1}{2}$  26.  $\frac{1}{2}$  27.  $\frac{1}{2}$  28.  $\frac{1}{2}$  29.  $\frac{1}{2}$  30.  $\frac{1}{2}$  31.  $\frac{1}{2}$  32.  $\frac{1}{2}$  33.  $\frac{1}{2}$  34.  $\frac{1}{2}$  35.  $\frac{1}{2}$  36.  $\frac{1}{2}$  37.  $\frac{1}{2}$  38.  $\frac{1}{2}$  39.  $\frac{1}{2}$  40.  $\frac{1}{2}$  41.  $\frac{1}{2}$  42.  $\frac{1}{2}$  43.  $\frac{1}{2}$  44.  $\frac{1}{2}$  45.  $\frac{1}{2}$  46.  $\frac{1}{2}$  47.  $\frac{1}{2}$  48.  $\frac{1}{2}$  49.  $\frac{1}{2}$  50.  $\frac{1}{2}$  51.  $\frac{1}{2}$  52.  $\frac{1}{2}$  53.  $\frac{1}{2}$  54.  $\frac{1}{2}$  55.  $\frac{1}{2}$  56.  $\frac{1}{2}$  57.  $\frac{1}{2}$  58.  $\frac{1}{2}$  59.  $\frac{1}{2}$  60.  $\frac{1}{2}$  61.  $\frac{1}{2}$  62.  $\frac{1}{2}$  63.  $\frac{1}{2}$  64.  $\frac{1}{2}$  65.  $\frac{1}{2}$  66.  $\frac{1}{2}$  67.  $\frac{1}{2}$  68.  $\frac{1}{2}$  69.  $\frac{1}{2}$  70.  $\frac{1}{2}$  71.  $\frac{1}{2}$  72.  $\frac{1}{2}$  73.  $\frac{1}{2}$  74.  $\frac{1}{2}$  75.  $\frac{1}{2}$  76.  $\frac{1}{2}$  77.  $\frac{1}{2}$  78.  $\frac{1}{2}$  79.  $\frac{1}{2}$  80.  $\frac{1}{2}$  81.  $\frac{1}{2}$  82.  $\frac{1}{2}$  83.  $\frac{1}{2}$  84.  $\frac{1}{2}$  85.  $\frac{1}{2}$  86.  $\frac{1}{2}$  87.  $\frac{1}{2}$  88.  $\frac{1}{2}$  89.  $\frac{1}{2}$  90.  $\frac{1}{2}$  91.  $\frac{1}{2}$  92.  $\frac{1}{2}$  93.  $\frac{1}{2}$  94.  $\frac{1}{2}$  95.  $\frac{1}{2}$  96.  $\frac{1}{2}$  97.  $\frac{1}{2}$  98.  $\frac{1}{2}$  99.  $\frac{1}{2}$  100.  $\frac{1}{2}$

087  
-----  
| | | | | : |  
| | | | | : |

201 -----WKRRPLPKLMEKVLKRAKALPL 313

DECREE 10

XX

XX  
-----  
(continued)

adenine glycosylase; muty; Vaccination; immune response; gastritis; H pylori infection; stomach cancer

Streptococcus pneumoniae.

PD 06-MAY-1999.

27-OCT-1997: 097HS-0058676

XX  
T





OY 267 -----GEVHALTHRLRAVE-VRGAL-----N-----EGEGEDPKRPLPKL 302  
DB 315 PQWTDNHFNIKHTFSHOKWTELEEGYVAKATDLRNAPHLKMKVALEDFSLYFPAIRPOKKM 374  
OY 303 MEKVLAK 309  
DB 375 LETYLLKQ 381

RESULT 12

ID AAM60244 standard; Protein: 350 AA.

AA60244;

19-AUG-1998 (first entry)

Amino acid of the specification.

Human; MYH; bMYH; diagnosis; cancer.

Unidentified.

JPI0057076-A.

03-MAR-1998.

11-MAR-1997; 97JP-0099540.

11-MAR-1996; 96US-0013132.

(HUMA-) HUMAN GENOME SCI INC.

WPI: 1998-210407/19.

New isolated polypeptide - useful for, e.g. diagnosis of cancer

Disclosure; Page 22; 23pp; Japanese.

The present sequence appears in the specification. The specification describes a human MYH (bMYH) cDNA and protein. The MYH polypeptide can be used for diagnosing cancer, where the decreasing levels of the polypeptide are measured in a sample from a patient.

Sequence 350 AA:

Query Match 27.3%; Score 464.5; DB 19; Length 350;

Best Local Similarity 34.8%; Pred. No. 1.1e-35;

Matches 120; Conservative 61; Mismatches 121; Indels 43; Gaps 13;

OY 8 LLAATREMAR-PLPWRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66  
DB 10 VLDMWDKGRKTLPMQIDKTPYKVLSEMLQOTVAIVYIFEFERFMAFPVITDLANAP 69  
OY 67 LEEVLVWOGAGYVRAEHLRLARSVEE-----PPSFAELRGJLPGIGPYTAAVASTA 121  
DB 70 LDEVLHMTGLGTVARARLHKAQOVATLHGKPFPEFEVVAALPGVGRSTAGAILLS 129  
OY 122 FGERAAADGNVRRVLSLRFAPES-PKEKE-----LFALAQGLPE-GVDPGVWNOALMEL 175  
DB 130 LKHNPIILDGNKRVLAFCYAVSGWPKRKEVENKLMSLEQYTPAVGVER--FNQAMMDL 187  
OY 176 GATVCLPKRPRGACPL--GAFCRGKEAPRGYRPARKRRAKERLVALVLLGRKGVHL- 231  
DB 188 GAMITRTRSKPKSCSLCPDLONGCILAANNNSWALYPGKKPKQOTLPERKGTFLLLQHEDEVLLA 247  
OY 232 -ERLEGRFOGLYVLPFPEE-----LPGREAAGFVRSRPLGVEVNHATLHRLRAY----- 280  
DB 248 QRPESGLMGILYCFQFADDESLRQWLAORQIAANLNO-LTAFRHITSHPILDLVPMML 306  
OY 261 -----EVRGALMEGBGDEP---WKRPLFKLMEKVLKRALPL 313

DB 307 PYVSEFTGCMDEGNALMYLAQPPSVGLAAPERLLQO-LRTGAPV 350

RESULT 13

ABP30125

ID ABP30125 standard; Protein: 374 AA.

ABP30125;

02-JUL-2002 (first entry)

Streptococcus polypeptide SEQ ID NO 9426.

Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

group A streptococcus; Streptococcus pyogenes; antibacterial;

antiinflammatory; infection; vaccine; meningitis; gene therapy.

Streptococcus agalactiae.

WO200234771-A2.

02-MAY-2002.

29-OCT-2001; 2001WO-GB04789.

27-OCT-2000; 2000GB-0026333.

24-NOV-2000; 2000GB-0028727.

07-MAR-2001; 2001GB-0005640.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

Tettelin H;

WPI: 2002-352536/38.

N-PSDB: ABN70756.

New Streptococcus protein for the treatment or prevention of infection

or disease caused by Streptococcus bacteria, such as meningitis, and

for detecting a compound that binds to the protein -

Claim 1; Page 4069; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B

Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

(Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in

the specification. The proteins have antibacterial and antiinflammatory

activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and

antibodies that bind (I) are used in the manufacture of medicaments for

the treatment or prevention of infection or disease caused by

Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

Nucleic acids encoding (I) are used to detect Streptococcus in a

biological sample. (I) is used to determine whether a compound binds to

(I). A composition comprising (I) or a nucleic acid encoding (I), may be

used as a vaccine or diagnostic composition. The disease caused by

Streptococcus that is prevented or treated may be meningitis. Nucleic

acid encoding (I) may be used to recombinantly produce (I) and may be

used in gene therapy. Antibodies to (I) are used for affinity

chromatography, immunoassays, and distinguishing/identifying

Streptococcus proteins.

Sequence 374 AA:

Query Match 27.1%; Score 461; DB 23; Length 374;

Best Local Similarity 31.1%; Pred. No. 2.7e-35;

Matches 114; Conservative 68; Mismatches 125; Indels 60; Gaps 12;

OY 1 VEAMRKALAWYREMARPLPWRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLK 60  
DB 7 IASFRKTLGMYDCKRKLPMKRRTTNPIYVWSEMLQOTVAVIPIYKRFLEMFQIK 66  
OY 61 ALAASLEBVLVWOGAGYVRAEHLRLARSVEE-----LPPSFAELRGJLPGIPYTA 115

```

Db      67 DLADAEDEDLKAMEGLGYSSRVNRNMQAKAQVWVDGGLFPHTYDDIASLKGIPYTAG 126
      116 AVASINAFGERVAADVGNRVLSRLPARE-----SPEKEELFALAOGLLPBGVDPCWNOA 171
      127 AIASISFNNLPPEPVADNVMARLFEVNDIDDPNKRKLFQAIMELIDPDRGDPNOA 186
      172 LMELGATVCLLPKPRGACPL-----GAFRCGEAPGRYPAPR-KRAKEERLVALVLTGR 226
      187 LMDLGIDIESAKTIPRDESPTRFNNAYLNGTYS--KTPYKNTKKPKPRIOQFVLRNO 244
      227 KGVHL--ERLEGR-FQGLYGVPLFPPEELPGREAPF-GVRSRPL----- 266
      245 NGQYLLEKNTKGRLLGFMFSFPIETSPLSQQLDFDDNQNPIIMQTONETFOREYQTK 304
      267 -----GEVRALTHRLRLRV-VGAL-----W---EGEGDDPKRPLPKL 302
      305 PONTDNHFPNPKHTFESHOKWTIELIGVVKATDLPNAPHLKVAIEDFSLYPFPATPQKM 364
      303 MEKVLKR 309
      365 LETYTKQ 371

```

RESULT 14  
ID AAY37104 standard; Protein: 341 AA.

AC AAY37104;  
DT 07-OCT-1999 (first entry)

Protein involved in intermediate metabolism of nucleic acids.

Vaccine: eye disease; conventional trachoma; nonendemic trachoma;  
paratrachoma; inclusion conjunctivitis; genital disease; peritphalitis;  
nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
bartholinitis; pneumopathy; venereal lymphogranulomatosis.

Chlamydia trachomatis.

W09928475-A2.

10-JUN-1999.

27-NOV-1998: 98WO-IB01939.

04-NOV-1998: 98US-010707.

28-NOV-1997: 97FR-0015041.

17-DEC-1997: 97FR-0016034.

(GEST ) GENSET.

Griffals R;

WPI; 1999-371125/31.

Genome sequence of Chlamydia trachomatis

Disclosure: Page 902; 1755pp; English.

AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome of Chlamydia trachomatis (see AA201425). The polypeptides can be used as vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences can also be used to control growth of the microorganism. Chlamydia trachomatis is responsible for a large number of diseases, e.g. eye diseases such as conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion conjunctivitis; genital diseases such as nongonococcal urethritis, epididymitis, cervicitis, salpingitis, peritphalitis, bartholinitis; pneumopathy in breast feeding infants; and venereal lymphogranulomatosis. The polypeptides of the invention may be of use in treating these diseases.

Sequence 341 AA;

Query Match 27.1%; Score 460.5; DB: 20; Length 341;  
Best Local Similarity 39.1%; Pred. No. 2.7e-35;  
Matches 111; Conservative 49; Mismatches 103; Indels 21; Gaps 9;

```

      6 KALLAWRENARPLPMGKEDPYRVLYSEVLLQOTRYEQLAPYRRFLERFPTLKALAA 65
      26 EALRSWPLESKRSRPMDSPTPRVWVSEWMLQOTRAEVVYPTFLKMEFRPTLODLAQA 85
      66 SLEEVLRWQAGAGYRRARHLRLARSVE-----ELPPSAELRGILGEPYTAANAASI 120
      86 RESDVQWMEGLGYSSRVNRLAGARVITELFGEIYNDLALLSIGISYANALIAF 145
      121 AFGERVAADVGNRVLSRLPARESP-----KEKEELFALAOGLLPBGVDPCWNOALMEL 175
      146 AFQOKNPADVGNVLRVMSRLFAIESIDRNMTREITGLCESLLPD-QDPQVIAESFIEL 204
      176 GATVCLPKPRGACPLGAFRCRG--KEAPGRYPAPRKRakeERL--VALVLGRKGVHL 231
      205 GARIC-KQPLCECPURSFCTAYRGSTMEQIPV-RNTPRAISRLFRAVYIVLYKQDYLM 262
      232 ERLEGR--FQGLYGVPLF--PPEELPGREAPFGRSRPLGEVRH 271
      263 TKREKEIMAGLYEFPYQLPKEDCDIEKIHLYQKDYGETLH 306

```

RESULT 15  
ID ABU01657 standard; Protein: 391 AA.

AC ABU01657;

DT 11-FEB-2003 (first entry)

S. pneumoniae type 4 strain protein from coding region #1233.

Bacterial meningitis; pneumonia; sepsis; otitis media;

ear infection; antiinflammatory; antibacterial; immunostimulant;

auditory; respiratory; gene therapy; vaccine.

Streptococcus pneumoniae type 4 strain.

W0200277021-A2.

03-OCT-2002.

27-MAR-2002; 2002WO-IB02163.

27-MAR-2001; 2001GB-0007658.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Maignani V, Tettelin H, Fraser C;

WPI; 2003-040579/03.

N-PDB; ABX06945.

New proteins and nucleic acid molecules from Streptococcus pneumoniae,

useful as medicaments for treating or preventing a disease or infection

due to streptococcus bacteria, such as pneumonia, sepsis, otitis media

or ear infection

Claim 1: SEQ ID No 2466; 56pp; English.

The invention relates to a protein comprising or having at least 50%

identity to any of the 2469 amino acid sequences, identified in the

specificity (available on a computer readable format), or its fragment,

expressed from 2469 of 2489 identified DNA coding regions from the

Streptococcus pneumoniae type 4 strain genomic sequence appearing as

AB556454. Also included are an antibody which binds one of the

proteins, treating a patient by administering the protein, DNA or

CC antibody in a composition), a kit comprising first and second primers,  
CC which are the nucleic acid cited above or fragments between nucleotides  
CC 8-100 of a sequence not defined in the specification, for amplifying a  
CC target sequence contained within a Streptococcus nucleic acid sequence,  
CC where the first primer is substantially complementary to the target  
CC sequence and the second primer is substantially complementary to the  
CC complement of the target sequence, and where the parts of the primers  
CC having substantial complementarity define the termini of the target  
CC sequence to be amplified, assay comprising contacting a test compound  
CC with the protein, and determining whether the test compound binds to the  
CC protein and a Streptococcus pneumoniae bacterium, where one or more  
CC genes encoding the proteins has been rendered inactive. The proteins,  
CC nucleic acid molecules, antibody and compositions are useful as  
CC medicaments for treating or preventing a disease or infection due to  
CC streptococcus bacteria, particularly *S. pneumoniae*, such as pneumonia,  
CC sepsis, otitis media or ear infection. They are also useful in developing  
CC vaccines, diagnostics and antibiotics. The methods are useful for  
CC identifying immunodominant proteins. The present sequence is one of  
CC the 2469 proteins expressed by the identified coding regions from the  
CC genomic sequence.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences.

**SQ Sequence 391 AA;**

Query Match	27.08;	Score 459.5;	DB 24;	Length 391;
-------------	--------	--------------	--------	-------------

Best Local Similarity 33.98; Pred. NO. 4e-35;  
Matches 123; Conservative 53; Mismatches 128; Indels 59; Gaps 11.

[illegible]

Search completed: September 15, 2003, 13:50:45  
Job time : 86 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: September 15, 2003, 13:47:45 ; Search time 97 Seconds  
(without alignments)  
864.609 Million cell updates/sec

Title: US-09-938-901-2

Perfect score: 1700  
Sequence: 1 YEAMRKALAWYRENAERPLP.....YLRKALPLAHAGVVPDPDA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*

- 1: sp-archaea:\*
- 2: sp-bacteria:\*
- 3: sp-fungi:\*
- 4: sp-human:\*
- 5: sp-invertebrate:\*
- 6: sp-mammal:\*
- 7: sp-mhc:\*
- 8: sp-organelle:\*
- 9: sp-phage:\*
- 10: sp-plant:\*
- 11: sp-rodent:\*
- 12: sp-virus:\*
- 13: sp-vertebrate:\*
- 14: sp-unclassified:\*
- 15: sp-virus:\*
- 16: sp-bacteriap:\*
- 17: sp-archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	669.5	39.4	363	16	Q9RS42
2	552.5	32.5	358	16	Q8G240
3	551.5	32.4	375	16	Q8YFS8
4	531	31.2	366	16	Q92RF0
5	520	30.6	349	16	Q9AB55
6	515.5	30.3	367	16	Q8UH85
7	514.5	30.3	396	16	Q985U7
8	505	29.7	368	16	Q8DJZ3
9	503	29.6	349	16	Q9JTV7
10	501.5	29.5	369	16	Q31584
11	498	29.3	346	2	Q9ZEF1
12	498	29.3	349	16	Q9JYW9
13	497.5	29.3	372	16	Q9KEC2
14	492.5	29.0	354	16	Q8CV61
15	492	28.9	387	16	Q9CH97
16	491.5	28.9	350	16	Q8Z3U0

17	491	28.9	371	16	Q8ZHE8	Q8ZHE8 yersinia pe
18	491	28.9	415	16	Q8CZV0	Q8CZV0 yersinia pe
19	484.5	28.5	362	16	Q8Y2D5	Q8Y2D5 ralsstia s
20	484.5	28.5	521	4	Q9UBP2	Q9UBP2 homo sapien
21	484.5	28.5	522	4	Q9UBP4	Q9UBP4 homo sapien
22	484.5	28.5	532	4	Q9UF5	Q9UF5 homo sapien
23	484.5	28.5	535	4	Q15830	Q15830 homo sapien
24	484.5	28.5	536	4	Q9UIF6	Q9UIF6 homo sapien
25	484.5	28.5	546	4	Q9UIF7	Q9UIF7 homo sapien
26	483	28.4	365	16	Q9ZAX1	Q9ZAX1 listeria in
27	478.5	28.1	353	16	Q9KUR3	Q9KUR3 vibrio chol
28	478	28.1	355	16	Q8Y6J5	Q8Y6J5 listeria mo
29	477.5	28.1	355	16	Q9HU37	Q9HU37 pseudomonas
30	477	28.1	369	16	Q9Z8E1	Q9Z8E1 chlamydia p
31	476	28.0	339	16	Q9PM55	Q9PM55 campylobact
32	472.5	27.8	365	16	Q8EBX7	Q8EBX7 shewanella
33	470.5	27.7	350	16	Q8FE21	Q8FE21 escherichia
34	469.5	27.6	345	16	Q8XCS8	Q8XCS8 escherichia
35	468.5	27.6	345	16	Q8DSD4	Q8DSD4 streptococc
36	467	27.5	345	16	Q99T11	Q99T11 staphylococ
37	466	27.4	374	16	Q8N2I9	Q8N2I9 streptococc
38	466	27.4	384	16	Q8K5Y9	Q8K5Y9 streptococc
39	465	27.4	345	16	Q8NVU5	Q8NVU5 staphylococ
40	465	27.4	374	16	Q99Y77	Q99Y77 streptococc
41	462	27.2	374	16	Q8E3J9	Q8E3J9 streptococc
42	461	27.1	374	16	Q8DXV0	Q8DXV0 streptococc
43	461	27.1	515	11	Q99P21	Q99P21 mus muscull
44	460.5	27.1	516	11	Q8R5G2	Q8R5G2 rattus norv
45	459.5	27.0	361	16	Q97Q13	Q97Q13 streptococc

## ALIGNMENTS

RESULT 1

Q9RS42 PRELIMINARY; PRT; 363 AA.

AC Q9RS42;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 GN A/G-specific adenine glycosylase.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
 OC Deinococcaceae; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RI;  
 RA MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hikey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat R.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 RT radiodurans RI.";  
 RL Science 286:1571-1577(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RI;  
 RA MEDLINE=21475757; PubMed=11591657;  
 RA Li X., Lu A.L.;  
 RT "Molecular Cloning and Functional Analysis of the MutY Homolog of  
 RT Deinococcus radiodurans";  
 RL J. Bacteriol. 183:6151-6158(2001).  
 DR EMBL: AE002060; AAF11831.1; -;  
 DR EMBL: AF377342; AAL26976.1; -;  
 DR HSP: P17802; IMUN.  
 TR: TIGR: DR2285; -;

DR InterPro: IPR004035; EndoIII\_FCL.  
DR InterPro: IPR003265; Endo\_3c.  
DR InterPro: IPR003651; Res.Dind.  
DR InterPro: IPR005760; Mult.  
DR InterPro: IPR001680; WD40.  
DR Pfam: PF00730; Hnh-GPD; 1.  
DR SMART: SM00478; ENDO3c; 1.  
DR SMART: SM00525; FES; 1.  
DR TIGRFAMs: TIGR01084; muty; 1.  
DR PROSITE: PS00764; ENDONUCLEASE\_III\_1; 1.  
DR PROSITE: PS00678; WD\_REPEATS\_1; 1.  
DR Complete proteome.  
KW SEQUENCE 363 AA; 39388 MW; 1D45C6AA09A2EE2F CRC64;

Query Match 39.4%; Score 669.5; DB 16; Length 363;  
Best Local Similarity 46.2%; Pred. No. 4.7e-42;  
Matches 154; Conservative 48; Mismatches 96; Indels 35; Gaps 11;

QY 1 VEAMRKALLMYRENAPRLPWR-----GEKDPYRVLVSEVLLQOTRVGEQALPYRRFLERF 56  
DB 17 VGALRRDLGWDFRAGNDLPWRLGDEGRDPYRVVVAEILLQOTVAVAGLGYERFLEAF 76  
QY 57 PTLKALAAASLEEVLRVWQAGYTRAEHLRLARSVE--LPPSFAELRGLPGLPYTA 114  
DB 77 PTVOALAAAPQDAVLKAWEGGYYARARNLHRAAIIDEQGFPODYAGMLALPGVPYTA 136  
QY 115 AAASIAFGERVAADVGNVRVLSRLFARESPKKEKELFALQGLPEGVDPGVNQAALME 174  
DB 137 AAASISLALGERVAADVGNVRVLSRLFARESPKKEKELFALQGLPEGVDPGVNQAALME 195  
QY 175 LGATVCLPKRRPGCAGPLAFCKGKE--APGRYPAPRRK-RAKEERLVALVLLGRKKVHL 231  
DB 196 LGATICVPKSPACRCFVSAHCAVYQLQPGDPFAPKARPOAREVRAVALLLIGDAEYAVL 255  
QY 232 ERLEGR-FQGIYGVPLPEPEELPGRFAFGVRSR-----PLGEVHALTHRLR 279  
DB 256 EKRGSLGLFGLPL--EIGARETADALARLQARLGAKEVKECTGVGHMTHRLS 312  
QY 280 VEVRGALMEGEGEDPMKRP-----LPKLMKRYL 307  
DB 313 VEV-----YRAEADRP-RQPVRGALSLRDHKL 340

RESULT 2  
Q8G240 PRELIMINARY; PRT; 358 AA.  
AC O8G240;  
DT 01-MAR-2003 (TREMblrel. 23, Created)  
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
DE A/G-specific adenine glycosylase.  
GN Muty OR BR0493.  
OS Brucella suis.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Brucellaceae; Brucella.  
OX NCBI\_TaxID=29461;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1330 / Biovar 1;  
RA MEDLINE=22247741; Pubmed=12271122;  
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,  
RA Read T.D., Dodson R.J., Unayam L., Brinkac L.M., Beanan M.J.,  
RA Daugherty S.C., Debey R.T., Durkin A.S., Kolonay J.F., Madupu R.,  
RA Nelson W.C., Ayelel B., Krahl S., Shetty J., Malik J., Van Aken S.E.,  
RA Riedmiller S., Jettelin H., Gill S.R., Boyle O., Salzberg S.L.,  
RA Hoover D.L., Lindler L.E., Halling S.W., White C., Salzberg S.L.,  
RT "The Brucella suis genome reveals fundamental similarities between  
RT animal and plant pathogens and symbionts."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).  
DR EMBL: AE014358; AAN29436.1; -  
DR TIGR: BR0493; -  
KW Complete proteome.  
SQ SEQUENCE 358 AA; 39047 MW; E1478C3CEDB74E7 CRC64;

Query Match 32.5%; Score 552.5; DB 16; Length 358;  
Best Local Similarity 38.4%; Pred. No. 2.4e-33;  
Matches 134; Conservative 53; Mismatches 111; Indels 51; Gaps 12;

QY 8 LLAWYRENAPRLPWR-----GE-KDPYRVLVSEVLLQOTRVGEQALPYRRFLERPT 58  
DB 7 LLRWYDRHHRVLPWRVPPVDAKGDVADPVVWVLSIMLQOTVAAKSVFLREIEMPT 66  
QY 59 LKALAAASLEEVLRVWQAGYTRAEHLRLARSV-----ELPPSFAELRGLPGLPYTA 113  
DB 67 VRAAKASEDDILKAWAGLGYSRARNLKCADIVVAEHGGEFPAKSAAGLKELPGIGDT 126  
QY 114 AAASIAFGERVAADVGNVRVLSRLFARESPKKEKELFALQGLPEGVDPGVNQAALME 170  
DB 127 SAATAIAIAFGQVAVNDGNVRVLSRLFAIDPLPVAKAQICALMGMTPTD-RPGPQAQ 185  
QY 171 ALMELGATVCLPKRRPGCAGPLAFCKG--KEAQRIP--APRRRAKEERLVALVLLGR 226  
DB 186 AMMDLGATICTPRRPAACALCPLNKGCIALCERDEDEFPVKAPEKPYRTGAFAIAAGD 245  
QY 227 KGVHLERLEGRFQGIYGVPLPEPELP9-----REAAGVRSRPLGEVRHAL 273  
DB 246 GSVTLRRKKG--EGL---LAGMTEVPGSGTARIDGATVNAAPFSAAWTPTSGTTHVF 299  
QY 274 THRLRVEVRGA---LWEGEGEDPMKRP-----LPKLMKRYLRKALP 312  
DB 300 THFELRLSVYRASNVKQANEGWMSPEELCGEALPTVMKKALIAAIP 348

RESULT 3  
Q8YFS8 PRELIMINARY; PRT; 375 AA.  
AC Q8YFS8;  
DT 01-MAR-2002 (TREMblrel. 20, Created)  
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
DE A/G-specific adenine glycosylase (EC 3.2.2.-).  
GN BMEI1442.  
OS Brucella melitensis.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Brucellaceae; Brucella.  
OX NCBI\_TaxID=29459;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=16M / ATCC 23456 / Biotype 1;  
RA MEDLINE=20020109; Pubmed=11756688;  
RA Delvecchio V.G., Kapatal V., Redkar R.J., Patra G., Mufier C., Los T.,  
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,  
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldman E.,  
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Teleson J.-J.,  
RA Haselkorn R., Kyrides N., Overbeek R.;  
RT "The genome sequence of the facultative intracellular pathogen  
RT Brucella melitensis."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
DR EMBL: AE009581; AAL52623.1; -  
DR InterPro: IPR004035; EndoIII\_FCL.  
DR InterPro: IPR003265; Endo\_3c.  
DR InterPro: IPR003651; Res.Dind.  
DR InterPro: IPR005760; Mult.  
DR Pfam: PF00730; Hnh-GPD; 1.  
DR SMART: SM00478; ENDO3c; 1.  
DR SMART: SM00525; FES; 1.  
DR TIGRFAMs: TIGR01084; muty; 1.  
DR PROSITE: PS00764; ENDONUCLEASE\_III\_1; 1.  
KW Hydrolase; Glycosidase; Complete proteome.  
SQ SEQUENCE 375 AA; 40938 MW; 2E5FD5E76533B46D CRC64;

Query Match 32.4%; Score 551.5; DB 16; Length 375;  
Best Local Similarity 38.4%; Pred. No. 3.1e-33;  
Matches 134; Conservative 53; Mismatches 111; Indels 51; Gaps 12;  
QY 8 LLAWYRENAPRLPWR-----GE-KDPYRVLVSEVLLQOTRVGEQALPYRRFLERPT 58

Db 24 LRRYDRHHHVLPRVTPVDAKGDVADPYKRWLSEIMLQOTVEAVKSYLRFIERMPT 83  
QY 59 LKALAAASLEEVLRVWQAGCYRRAEHLRLARSV-----ELPPSPALGLGLGYT 113  
Db 84 VRAAKASSEDDILKAMAGLGYSSRAHLKCKADIVVAHEGGEFPRKSAAGLELPGIGYT 143  
QY 114 AAASIAFGERVAVAGNVRVLSRLFARES- -KEKEFLAAGLPGVPGVWQ 170  
Db 144 SAATAALAFGEQVAVVGNVERVSRILXALDTPLPVAKAQIRALMGQTPPD- RPDGPAQ 202  
QY 171 ALMEIGATVCLPKRPGACPLGACRG- -KEARGRP-APRRRAKEERLVALVLGR 226  
Db 203 AMMDLGATICTPRRACALCP LNKGCIALCERODEPPVAKPKAEKPYRTGAFIALAGD 262  
QY 227 KGVHLERLEGRFQGLVGPFPPEELPG-----REAFGVRSRPLGEVRAL 273  
Db 263 GSVYLRKRKG--EGL---LAGMTEVPGSGMTARIDDAIVNAAPFSAAMTPSGTITHVF 316  
QY 274 THRRLRYEVGA---LWEGEGEDPKRP-----LPKIMEKVKRLP 312  
Db 317 THEFLRLSYRASVNRKQANEGWSTPEELGALPTVMKKAIAAIP 365

## RESULT 4

Q92RF0 PRELIMINARY; PRT; 366 AA.  
AC Q92RF0; 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Probable A/G-specific adenine glycosylase protein.  
GN MMY OR R00928 OR SMC00452.  
OS Rhizobium melioli (Sinorhizobium melioli).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021.  
RX Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
RA Godrie T., Goffeau A., Kahn D., Kise E., Lelaure V., Masuy D.,  
RA Pohl T., Portetle D., Puehler A., Purnelle B., Ramspeger U.,  
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.,  
RT "Analysis of the chromosome sequence of the legume symbiont  
RT Sinorhizobium melioli strain 1021."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
DR EMBL: AL591785; CAC45500.1; -.  
DR InterPro: IPR004035; EndoIII\_FCL.  
DR InterPro: IPR003265; Endo\_3c.  
DR InterPro: IPR003651; Fes\_Bind.  
DR InterPro: IPR000445; Hnh.  
DR InterPro: IPR005760; MutY.  
DR Pfam: PF00633; HNH.1.  
DR Pfam: PF00730; Hnh-GPD.1.  
DR SMART: SM00478; ENDO3c.1.  
DR SMART: SM00525; FES.1.  
DR TIGRFAMS: TIGR01084; muty.1.  
DR PROSITE: PS00764; ENDONUCLEASE\_III\_1.1.  
KM Complete proteome.  
SQ SEQUENCE 366 AA; 39944 MW; 698224AA25089831 CRC64;

## Query Match

Best Local Similarity 31.2%; Score 531; DB 16; Length 366;  
Matches 136; Conservative 41; Mismatches 125; Indels 60; Gaps 11;

QY 8 LLAMYENARPLPWR-----GKDPYRVLSVYLLOOTRVEQALPYRRLEPPT 58  
Db 14 LLEMYDRHHHDLPRVTPVDAKGDVADPYKRWLSEIMLQOTVEAVKSYLRFIERMPT 73  
QY 59 LKALAAASLEEVLRVWQAGCYRRAEHLRLARSV-----ELPPSPALGLGLGYT 113

Db 74 VGDLAADTEEDVVMKAMAGLGYRARRNLKCAEAVARDBGRFPDSEEGALPGIGYT 133  
QY 114 AAASIAFGERVAVAGNVRVLSRLFARES- -KEKEFLAAGLPGVPGVWQ 170  
Db 134 AAATAALAFNPNASVLDGNGVERVSRILXAVETPLPAKPKAEKPYRTGAFIALAGD 192  
QY 171 ALMEIGATVCLPKRPGACPLGACRG- -KEARGRP-APRRRAKEERLVALVLGR 230  
Db 193 AMMDLGATICTPRRACALCP LNKGCIALCERODEPPVAKPKAEKPYRTGAFIALAGD 248  
QY 231 LERLEGRF-----QGLYGVPLFPPEELPGRE-----AAGVRSRPLGEVRH 271  
Db 249 VDGLEAVYLRKRPEFTGLG-----GMTEVPGTDMVSRDGTSDAHPPAEPECGTYNH 304  
QY 272 ALTHRLRYEV-----RGALWEGEGE-DPKRP-----LPKIMEKVKRLP 316  
Db 305 VFTHEFLRLSYRASVNRKQANEGWSTPEELGALPTVMKKAIAAIP 364  
QY 317 AG 318  
Db 365 AG 366

## RESULT 5

Q9AB55 PRELIMINARY; PRT; 349 AA.  
AC Q9AB55; 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE A/G-specific adenine glycosylase.  
GN CC0377.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;  
OC Caulobacteraceae; Caulobacter.  
OX NCBI\_TaxID=15892;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
RA Deboy R.T., Dodson R.J., Durkin A.S., Gilm M.L., Hatt D.H.,  
RA Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
RA Kletterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,  
RT "Complete genome sequence of Caulobacter crescentus."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL: AB005710; AKK2364.1; -.  
DR HSSP: P17802; IMON.  
DR TIGR: CC0377; -.  
DR InterPro: IPR003265; Endo\_3c.  
DR InterPro: IPR003651; Fes\_Bind.  
DR InterPro: IPR000445; Hnh.  
DR InterPro: IPR005760; MutY.  
DR Pfam: PF00633; HNH.1.  
DR Pfam: PF00730; Hnh-GPD.1.  
DR SMART: SM00478; ENDO3c.1.  
DR SMART: SM00525; FES.1.  
DR TIGRFAMS: TIGR01084; muty.1.  
KM Complete proteome.  
SQ SEQUENCE 349 AA; 37615 MW; DD96F16FB2A4ABC2 CRC64;

## Query Match

Best Local Similarity 30.6%; Score 520; DB 16; Length 349;  
Matches 141; Conservative 33; Mismatches 137; Indels 36; Gaps 10;

QY 2 EAWEKALLAMYENARPLPWR-----GKDPYRVLSVYLLOOTRVEQALPYRRLEPPT 52  
Db 5 DALRSALLAWYDAQARDLAWRVGPAERAGVRSPPYRWLSEIMLQOTVEAVKSYLRFIERMPT 64

```

OY 53 LEREPFLKALAAASLEEVLRWOGAGYVRAEHLRLARSVE-----LPPSFAELRGIP 107
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 65 TORMPVTDLAAVEDDILMAAAGLGYARARNTLACARAVANDGAVFPGTEGRLALP 124
OY 108 GLGPTTAAAVASIAIGERVAANDGVNRVLSRLFRRESP---KEKELFALAGLLEPGVD 164
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 125 VGAVATAAANAIAEDRAANVVDGNNVERMSRLFAVEAPMPDCKPELALADLTVD-D-R 183
OY 165 PGVMMQALMELGATVCLPKRPRCGACPLGAFCRG--KEAPGRYPAPRRKRAKEERL-VAL 221
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 184 PGDAQAOLMDLGATCTCKRKGRPLCDRCRPSLWCAATVGGARPEYPRKTKKADRRKHGAY 243
OY 222 VLLGRKGVHLERL--EGRFGGLGYVPL-----FPPEELPGREAAFGVSRPLGEVRR 272
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 244 VLTGDEVALVRRPRKGLLGMLGLPTSDMRTASTYDDAEAVA-AAPLAAAMRDLGAVEHV 302
OY 273 LTHRLRLREVGKALMEGEGEDPMK-----RPLPKLMEKVLKRALPL 314
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 303 FTHESLTLRVFAADGANDGDFVWTPREGIGALPVSFLKAAAMAQORLL 349

```

## RESULT 6

```

ID Q80H85 PRELIMINARY: PRT: 367 AA.
AC Q80H85;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
DE A/G-specific adenine glycosylase.
GN MUYT OR ATU0798 OR AGR_C_1460.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;

```

```

RP MEDLINE-21608550; PubMed=11743193;
RX MEDLINE-21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitaajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eilen J.A., Karp P.D., Boyce D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li W.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Seshphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chunley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Neester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quotillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Hummel K., Gordon J., Vaadin M., Iatchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Martelz B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Seer C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL: AE009046; AAL41814.1; -.
DR EMBL: AE008013; AAK86607.1; -.
DR InterPro: IPR004035; EndoIII_FCL.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR003651; Fes_bind.
DR InterPro: IPR005760; Muty.
DR Pfam: PF00730; Hnh-GPD; 1.
DR SMART: SM00478; ENDO3C; 1.
DR SMART: SM00525; FES; 1.
DR TIGRFAMS: TIGR01084; muty; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; 1.

```

```

KW Complete Proteome.
SQ SEQUENCE 367 AA; 40056 MW; 5237C00E3FF994A5 CRC64;

```

```

Query Match 30.3%; Score 515.5; DB 16; Length 367;
Best Local Similarity 37.1%; Pred. No. 1.4e-30;
Matches 130; Conservative 55; Mismatches 124; Indels 41; Gaps 13;

```

```

OY 3 AMRKALLAMYRENAARLPW-----RGKE-DPRVLVSEYLLLOOTNEQALPYRRRL 53
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 9 SYAQMLLAWYDRHNRRELPMRTSPMAARGRKADRPYVNMVLEWMLQOTTVQAVKPYELKFL 68
OY 54 ERFPTLKALAAASLEEVLRWOGAGYVRAEHLRLARSVE-----LPPSFAELRGIP 108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 69 AAMPVYDDLATPAPVEDVNAAMAGLGYARARLTKCAAVAREHGGVEPDTGEGKQLPG 128
OY 109 LGPYTAAAVASTAFGERVAANDGVNRVLSRLFRRESP---KEKELFALAGLLEPGVD 165
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 129 IGDYTSAAVAATAFNKQAAMVMDGNVERYSRLFALDAPLPGRKPMKAKVALLTP-AERP 187
OY 166 GVMNQALMELGATVCLPKRPRCGACPLGAFCRG--KEAPGRYPAPRRKRAKEERL-VAL 222
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 188 GDFQAQAMMDLGATCTCKRKGRPLCDRCRPSLWCAATVGGARPEYPRKTKKADRRKHGAY 247
OY 223 LIGRKGVLHLERL--EGRFGGLGYVPLP-PEELG-----RRAAGVSRPLGEVRRALT 274
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 248 AVNTGELLRLRRIDSGLGMTVEPTTAMTARMGTEVSHAPPAAGQMAGVIGHVFT 307
OY 275 HRRLRVEY-RGALMEG---EGEDPMKRP-----RPLMKELVKRALP 312
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 308 HFELRLTYRAQVDPGLGTGPDGMEPEVTNIDQAQLPVMKKVIAEALP 357

```

## RESULT 7

```

ID Q985U7 PRELIMINARY: PRT: 396 AA.
AC Q985U7;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)
DE Adenine glycosylase.
GN ML17523.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;

```

```

RP SEQUENCE FROM N.A.
RX STRAIN-MAFR303099;
RX MEDLINE-21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP003012; BAB53965.1; -.
DR InterPro: IPR004035; EndoIII_FCL.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR003651; Fes_bind.
DR InterPro: IPR005760; Muty.
DR Pfam: PF00730; Hnh-GPD; 1.
DR SMART: SM00478; ENDO3C; 1.
DR SMART: SM00525; FES; 1.
DR TIGRFAMS: TIGR01084; muty; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; 1.
KW Complete proteome.
SQ SEQUENCE 396 AA; 43649 MW; 2D3D11219667049E CRC64;

```

```

Query Match 30.3%; Score 514.5; DB 16; Length 396;
Best Local Similarity 37.6%; Pred. No. 1.9e-30;
Matches 130; Conservative 45; Mismatches 124; Indels 47; Gaps 10;

```



[illegible]

RESULT	ID	PRELIMINARY	PRT	368 AA.
08DUJ23	08DUJ23			
AC	08DUJ23			
DT	01-MAR-2003 (Tremblrel. 23, Created)			
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)			
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)			
DE	Adenine glycosylase.			
GN	TL1077.			
OS	Synechococcus elongatus (Thermosynechococcus elongatus).			
OC	Bacteria; Cyanobacteria; Chroococcales; Synechococcus.			
OX	NCBI_TaxID=32046;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SMRAIN-BP-1:			
RX	MEDLINE=2222514; PubMed=12240834;			
RA	Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,			
RA	Watanabe A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,			
RA	Kiyokawa C., Kohara C., Matsuno M., Matsuno A., Nakazaki N.,			
RA	Shimo S., Sugimoto M., Takeuchi C., Yamada M., Yabata S.,			
RT	"Complete genome structure of the thermophilic cyanobacterium			
RT	Thermosynechococcus elongatus BP-1."			
RL	DNA Res. 9:123-130(2002).			
DR	EMBL: AP005372; BAC00830.1; -.			
KW	Complete proteome.			
QO	SEQUENCE 368 AA; 41746 MW; CDE1BD4EB1A36143 CRC64;			

Query Match	29.7%;	Score 505;	DB 16;	Length 368;
Best Local Similarity	40.7%;	Pred. No. 8.8e-30;		
Matches 127;	Conservative 43;	Mismatches 102;	Indels 40;	Gaps 12

QY 5 AMKALLAAVREARLRLPRGKEDRYRVLYSVLLQOTRVQDALPYTRFLERPTLKAL 62  
15 ALFPALLNNVQOQGRDLPWRHRSRDPAWVISEIMLQOTVATVTPVYQRMVLTATPPTLBDL 74  
QY 63 AAASLEEVLRVMQAGCYRYRAELHRLRLARSV-----ELPSPFAELRCGLPGTGYTAAV 117  
75 AAAELETYVLTQMGLCYTARAHNLHRLRAQOIMTHINAGEFPSPSYEYVAVALPEIGSTGAT 134  
QY 118 ASIATSERVAAYDGNVRLVLSRLFARESPEKELEFAL---AQGLLEPGVDPGVWQAALM 173  
135 LSAFAFNOPDIPDLGNGKRVLARLYGLTVBPQKQAEQLQMWSAQQLLCPQ---SPRDFNOALM 192  
QY 174 ELGATVCLPRKPPRCGCPGLAFCRG-----KEAP---GRTPARKKRAKERLVALVLL 224  
Db 193 DLCAATICTPRHRCSTACSPWQHNCALHNRQLTHERIPKRSRSPPLPKKGG-----VAIWN 247

```
Oy      225 GRCGVHRL-EGRGQYXVP---LPEP---LPREP--AAFGVSRPLG-----VR 270
Db      248 ATGOILDRPPGLLGIMERPGGKIEPNETVOECIOREIREUGIEIR-VGEHLIDID 306
Oy      271 HALTHRRLEEV 282
Db      307 HAYTHEFVTLHV 318
```

RESULT	9
09JTV7	
ID	09JTV7
AC	09JTV7
DT	01-OCT-2000 (TREMBlrel. 15, Created)
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT	01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE	Adenine glycosylase (EC 3.2.2.-), MUTY OR NMA1614.
GN	
OS	Neisseria meningitidis (serogroup A).
OC	Bacteria: Proteobacteria: Betaproteobacteria; Neisseriales;
OC	Neisseriaceae; Neisseria.
OX	NCBI_taxid=65699;
	PRELIMINARY; PRT; 349 AA.

RP SEQUENCE FROM N.A.  
RC STRAIN-22491 / Serogroup A / serotype 4A:  
RX MEDLINE-20222556; PubMed-10761919;  
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
RA Davies R.M., Davis P., Devlin K., Fellwell T., Hamlin N., Holroyd S.,  
RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,  
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
RA Whitehead S., Spratt B.G., Barrrell B.G.;  
RT "Complete DNA sequence of a serogroup A strain of *Neisseria*  
RT meningitidis 22491.";  
RL Nature 404:502-506(2000).  
DR EMBL: AL162756; CAB8482.1; -.  
DR HSSP: P17802; 1MDN.  
DR InterPro: IPR004035; EndoIII\_FCU.  
DR InterPro: IPR003265; Endo\_3c.  
DR InterPro: IPR003651; Fes\_bind.  
DR InterPro: IPR000445; Hhh.  
DR InterPro: IPR005760; Muty.  
DR Pfam: PF00633; HHH\_1.  
DR Pfam: PF00730; Hhh-GPD; 1.  
DR SMART: SM00478; ENDO3c; 1.  
DR SMART: SM00525; FES; 1.  
DR TIGRfams: TIGR01084; muty; 1.  
DR PROSITE: PS00764; ENDONUCLEASE\_TII\_1; 1.  
KW Hydrolyase; Glycosylase; Complete proteome.  
SQ SEQUENCE 349 AA; 39560 MW; BBDEFF34613A1A9 CRC64;

Query Match	29.6%;	Score 503;	DB 16;	Length 349;
Best Local Similarity	39.7%;	Pred. No. 1.2e-29;		
Matches 131; Conservative	52;	Mismatches 111;	Indels 36;	Gaps 14

```

0Y      8  LLAMTRENAR-PLPMGGEKDPRLVLYSEVLLDQGTVEBALPYRRRELEPFTLKALAAAS 66
      14  LIRMOQHHGHLLPMO-VKNPYCVMWLSMTLMOOTVAALVDLYPPELEKFTVOTLAAAP 72
0Y      67  LEEVLRWMOGAGYTRAEHLHLRLARSVE-----LPSPFAELRGDLPGLPYTAAVAASIA 121
      73  ODEVLSTLWAGLGYTSARRLLHKAACOOVVEOFGFTPESEKRLDETLCGGRSTAAAIACFS 133
Db      122  FGEVRVAVDGNVRVRLFLARE-SPKEKE-----FPLAOGLLP-EGVDPCVMQALCMEI 175
0Y      133  ENRRETLIDGNVVRVLCRFARPDANGQDCKFENSLWTLAESELLSENNDMNATYQGLMDL 192
Db      176  GATVCLPKNRPGCAGCLGAFCKGKEAPNGYPPAPKRRKKE-----RLVALVYLGRKG-VHL 231
0Y      193  GATVCKRTRKGLCHQCEMADICEKKNONRTAELPRKKTAAEVPYLLPLWVLYLRNDDGAILL 252
0Y      232  E--LEGRFGGLGVLPPEELPGHE-----AAGVRSRPLGE-----VRHALLHRLRYEVR 283

```

Db 253 EKRPAKIGWGLVCVPCF--ESLNGLSDFAKFSLTWADMDQETALTHTLRLTLTPP 310  
 QY 284 GAWL--EGEGEDPWKRP-----LPKIME 304  
 Db 311 EAQWPSPSDGIMIKRAHLKDKGLRPLE 340

RESULT 10  
 031584  
 ID 031584 PRELIMINARY; PRT: 369 AA.  
 AC 031584;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE YFHQ protein.  
 GN YFHQ.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=938437;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borczyk S.,  
 Boriss R., Boursier L., Brans A., Brun M., Brignell S.C., Bron S.,  
 Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 Choi S.K., Codani J.Y., Conerton I.F., Cummings N.J., Daniel R.A.,  
 Deniot C.F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entlin K.D., Errington J., Fabret C., Ferrari E., Fougere D.,  
 Frits C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,  
 Guisepi G., Guy B.J., Hage K., Halech J., Harwood C.R., Henaut A.,  
 Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 Kobayashi Y., Koelter P., Konigstein G., Krogh S., Kumano M.,  
 Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
 Parro V., Pohl T.M., Portetlelle D., Porwollik S., Prescott A.M.,  
 Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,  
 Sato T., Scanlan E., Schleich S., Schnoeter R., Scoffone F.,  
 Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,  
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,  
 Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 "The complete genome sequence of the gram-positive bacterium Bacillus  
 subtilis".  
 RT Nature 390:249-256(1997).  
 RL Nature 390:249-256(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.,  
 RN Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97101647; PubMed=8946165;  
 RA Yamamoto H., Uchiyama S., Sekiguchi J.,  
 RT "Cloning and sequencing of a 27.8-kb nucleotide sequence of the 79  
 RT degrees-81 degrees region of the Bacillus subtilis genome containing  
 RT the sspE locus".  
 RL DNA Res. 3:257-262(1996).  
 DR EMBL: Z99108; CAB12691.1; -;  
 DR EMBL: D85082; BA24483.1; -;  
 DR HSSRP: P17802; IMUY.  
 DR InterPro: IPR003265; Endo\_3c.  
 DR InterPro: IPR003651; Fes\_bind.  
 DR InterPro: IPR000445; HHH.

DR InterPro: IPR005760; Muty.  
 DR Pfam: PR00633; HHH; 1.  
 DR Pfam: PF00730; HHH-GPD; 1.  
 DR SMART: SM00478; ENDO3c; 1.  
 DR SMART: SM00525; FES; 1.  
 DR TIGRFAMs: TIGR01084; mutY; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 369 AA; 41963 MW; FE1FE70018579B1 CRC64;

Query Match 29.5%; Score 501.5; DB 16; Length 369;  
 Best Local Similarity 36.1%; Pred. No. 1.6e-29;  
 Matches 112; Conservative 47; Mismatches 108; Indels 43; Gaps 6;

QY 1 VEAMRKALLAWRENAARPLPMWGEKDPKRVLYSEVTLQQTVEQALPYRRFLERPLPK 60  
 Db 13 IQQFDDLLSWFERQRLPWRDQDPYKVMSEVMLQOTREVEVYPLFLREVSQFPYVE 72  
 QY 61 ALAASLEBVLVWMOAGYVYRAEHLHRLASVEE----LPPSFAELRGLPGLYPTAA 115  
 Db 73 ALADADEKRVLLKAMEGGLGYSVNRNLSAIVEKQVEYCGIVPPDEKDEGLKGVPTTKG 132  
 QY 116 AVASTAFGEERVAADVGNRVLSRLFA---RESPEKEFLAAGLLPEGVDPGVNQA 171  
 Db 133 AVLSTAYNKKPIPAVDGVMYMSRLTSIMDDIAKPKRTTFEDAIRAFISKEKSPERNQG 192  
 QY 172 LMEIGATVCGLPKPRPGACPIGAFRCGRKEAPGRVAPAPKRAKE---ERLVATLLGRKG 228  
 Db 193 LMEIGALCTKRSPECCLCPVQHCNSAFEEGTRELPYKSKKKPGIKTMAVLTDEDG 252  
 QY 229 ---VHLERLEGRFOGLGVPLPPEELPGRBAEGVR-----SRP 265  
 Db 253 QVYIHKRPSKLLANLW-----EPNLETOGKIKTERBQLAFLENEYIQADISDL 304  
 QY 266 LGEVPHALTH 275  
 Db 305 QGVVEHVFTH 314

RESULT 11  
 092F41  
 ID 092F41 PRELIMINARY; PRT: 346 AA.  
 AC 092F41;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Adenine glycosylase.  
 GN Muty.  
 OS Neisseria meningitidis.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=487;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96118708; PubMed=7496539;  
 RA Kroll J.S., Langford P.R., Wilks K.E., Kell A.D.,  
 RT "Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct  
 RT from the eukaryotic enzyme, and not so rare after all".  
 RL Microbiology 141:2271-2279(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98084476; PubMed=9423860;  
 RA Wilks K.E., Dunn K.L., Farrant J.L., Reddin K.M., Corringe A.R.,  
 RL Langford P.R., Kroll J.S.,  
 RT "Periplasmic superoxide dismutase in meningococcal pathogenicity.";  
 RL Infect. Immun. 66:213-217(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98445380; PubMed=9770495;  
 RA Kroll J.S., Wilks K.E., Farrant J.L., Langford P.R.,  
 RT "Natural genetic exchange between Haemophilus and Neisseria:  
 RT intergeneric transfer of chromosomal genes between major human  
 RT pathogens".  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:12381-12385(1998).

DR EMBL; AJ001313; CAA04675.1; -.  
DR HSSP; PI7802; IMON.  
DR InterPro; IPR004035; EndoIII\_FCL.  
DR InterPro; IPR003265; Endo\_3c.  
DR InterPro; IPR003651; Fes\_bind.  
DR InterPro; IPR000445; Hnh.  
DR InterPro; IPR005760; Muty.  
DR Pfam; PF00663; HNH\_1.  
DR Pfam; PF00730; Hnh-GPD; 1.  
DR SMART; SM00478; ENDO3C; 1.  
DR SMART; SM00525; FES; 1.  
DR TIGRFAMs; TIGR01084; muty; 1.  
DR PROSITE; PS00764; ENDONUCLEASE\_III\_1; 1.  
DR SEQUENCE 346 AA; 39230 MW; 00AF2A26BEBEC3A1 CRC64;

Query Match	29.3%	Score 498	DB 2	Length 346
Best Local Similarly	39.4%	Pred. No. 2.7e-29		
Matches 130	Conservative 52	Mismatches 112	Indels 36	Gaps 14

[illegible]

RESULT 12  
Q9JYW9  
ID Q9JYW9 PRELIMINARY; PRT; 349 AA

DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE A/G-specific adenine glycosylase.  
GN NMB1396.  
OS *Neisseria meningitidis* (serogroup B).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC  
RC STRAIN=MC58 / Serogroup B;  
RX MEDLINE=20175755; PubMed=10710307;  
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
RA Eisten J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
RA Nelson W.C., Gilm M.L., Deboy R., Peterson J.D., Hickey E.K.,  
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
RA Mason T., Clecko A., Parisey D.S., Blair E., Clifton H., Clark E.B.,  
RA Colton M.D., Ulfelder T.R., Khoult H., Qin H., Vamathevan J.,  
RA Gill J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,  
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain  
RT MC58.";  
RL  
RL Science 287:1809-1815(2000).  
EMBL: AE002488; AAF41760.1; -

DR	HSSP: P17802.	-IMUN.
DR	TIGR: NMBJ396:	
DR	InterPro: IPR004035;	EndoIII_FCL.
DR	InterPro: IPR003265;	Endo_3c.
DR	InterPro: IPR003651;	Fes_bind.
DR	InterPro: IPR000445;	Hnh.
DR	InterPro: IPR005760;	Muty.
DR	Pfam: PF00633;	HHH; 1.
DR	Pfam: PF00730;	Hbh-GPD; 1.
DR	SMART: SM00478;	ENDOSC; 1.
DR	SMART: SM00525;	FES; 1.
DR	TIGRFAMs: TIGR01084;	muly; 1.
DR	PROSITE: PS00764;	ENDONUCLEASE_IIT_1; 1.
KW	Complete proteome.	
SQ	SEQUENCE	349 AA; 39587 MW; BB85FL34692AIA49 CRC64;
Query Match	29.3%;	Score 498; DB 16; Length 349;
Best Local Similarity	39.4%;	Pred. No. 2,7e-29;
Matches 130; Conservative	52;	Mismatches 112; Indels 36; Gaps 14

Query Match	Score	DB	Length
29.3%	498	16	
SEQUENCE	349 AA	39587 MW	BE85F134692A1A49 CRC64
SQ			

```

QY      8 LLAWREANAR--PLPRGEGDDPYRYVLSVEVLLLOOTREQALAPYRRLEPPRTKTLAAAS 66
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      14 LIRMOKQGRHNLPMQ--YKNPYCWLSEIMLOOTOYAIVLDIYPRLEKFPYQTLAAAP 72
QY      67 LEEVLRYWQAGACYRYRAEHLHRLARSVEE----LPPSFAELRGPLGDPYTAAVASIA 121
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      73 QDEVLSTWAGLGYTSRARNLHAAAOOVVQFGTFPSERKDDLETLGCVGSTAAALCAFS 132
QY      122 FGERVAADGANNRYVLSRLFARE--SPKEKE----LPLAQGLLP--EGVDPGVNQAALMEL 175
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      133 FNRRETILDGNNKRYVLCRFADGNNQDKKFEKSLMTLAESELLPENADMPAYTQGLMDL 192
QY      176 GATVCLPKRRPRCGACPLGAFCSGKEAPGRYPAPKRRAKEE---RLVALVLLGRKG--VHL 231
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      193 GATVCKRRTRPLCHOCPMADICCAKKONRTAEPLPRKKTAAEVPRLPYLWLVNRRNDGAILL 252
QY      232 ER--LEGRROGLGYAPLPPEELPGRE---AAFGVSRLPGE---VRNATLHRLREYVR 283
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      253 EKRPKKGIGWGLGYCPGF--ESLNGLSDEPAARFSLTMADMDEQALTHLRHLILLITPF 310
QY      284 GALW--EGEGEDDPMKRP-----LPKLTME 304
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      311 EAQMPSESPDGIWTKPAILKLDYQGLPKPLE 340

```

RESULT 13  
Q9KEC2  
ID Q9KEC2 PRELIMINARY; PRT; 372 AA

DT 01-OCT-2000 (TremBrel\_15, Created)  
DT 01-OCT-2000 (TremBrel\_15, Last sequence update)  
DT 01-MAR-2003 (TremBrel\_23, Last annotation update)  
DE Adenine glycosylase.  
GN BH0931.  
OS *Bacillus halodurans*.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.  
OX NCPL\_TaxID=6665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kihara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium *Bacillus*  
RT *halodurans* and genomic sequence comparison with *Bacillus subtilis*."  
RL Nucleic Acids Res. 28:4317-4333(2000).  
DR EMBL: AP001510; BAB04650.1; --  
DR HSSP: P17802; 1MU0.  
DR InterPro: IPR004035; EndoIII\_FCL.  
DR InterPro: IPR003265; Endo\_3c.  
DR InterPro: IPR003651; FeS\_bind.  
DR InterPro: IPR000445; Hhh.



QY 169 NOALMELGATVCLPKRRPGACPLGAFQ----RGKEAPGRYPAPRRR-RAKEERLYVALYL 223  
|||:|:| | | | | | :|  
Db 185 NOALMDLGSUVCSPKSPKCEACPLNYCAAAAGKQL--NYPVKTKKIKQKDLFTAFAL 242  
QY 224 LGRKG-VHLERLEGFRQGLYG-VPLFPPEELPGRE 256  
| :|:| :| | :| | | :|  
Db 243 ENSLGEYYLEKRPSK--GLADMTFPPLTELPAAD 275

Search completed: September 15, 2003, 13:53:00  
Job time : 101 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 13:40:49 ; Search time 23 Seconds

(without alignments)  
664,508 Million cell updates/sec

Title: US-09-938-901-2

Perfect score: 1700

Sequence: 1 YEAMRKALLAWYRENAAPLP.....VLKRALPLAHNGVPLPDA 325

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	484.5	28.5	350	1	MUTY_SALTY
2	471.5	27.7	350	1	MUTY_SALTY
3	445	26.2	378	1	MUTY_HAELIN
4	442.5	26.0	461	1	MYH_SCHPO
5	402	23.6	347	1	MUTY_BUCAP
6	378.5	22.3	350	1	MUTY_BUCAL
7	274.5	16.1	221	1	GMR_METFE
8	250.5	14.7	99	1	MUTY_AERHY
9	227	13.4	213	1	END3_THEMA
10	199	11.7	279	1	UVEN_MICLU
11	176	10.4	259	1	NTM1_CABEL
12	172	10.1	219	1	END3_BACSU
13	171	10.1	344	1	Y613_METVA
14	164.5	9.7	312	1	NTM1_HUMAN
15	161.5	9.5	300	1	NTM1_MOUSE
16	158.5	9.3	355	1	END3_SCHPO
17	158	9.3	219	1	END3_SYNY3
18	145	8.5	211	1	END3_SCOLI
19	134	7.9	211	1	END3_HAELIN
20	131.5	7.7	245	1	END3_MYCTU
21	130.5	7.7	245	1	END3_MYCLE
22	115.5	6.8	211	1	END3_TREPA
23	111.5	6.6	212	1	END3_TREPA
24	111	6.5	1729	1	END3_RICPR
25	107	6.3	1505	1	DME_ARATH
26	106	6.2	1309	1	CUT2_HUMAN
27	101.5	6.0	775	1	DMT2_ARATH
28	100.5	5.9	644	1	PMIP_SCHCO
29	100	5.9	1132	1	YAI4_TREPA
30	97.5	5.7	209	1	TERT_HUMAN
31	97.5	5.7	210	1	END3_BUCAP
32	97.5	5.7	432	1	END3_BUCAL
33	97	5.7	525	1	Y4IL_RHISN
					NAB2_HUMAN

34	97	5.7	635	1	VP40_HSV11
35	93	5.5	309	1	FDHE_PSEAE
36	93	5.5	1067	1	MICA_HUMAN
37	92.5	5.4	345	1	MUTY_THETH
38	92	5.4	502	1	GAG_SIVGB
39	91.5	5.4	376	1	PROB_THETH
40	90	5.3	873	1	FPS_FUSV
41	89.5	5.3	380	1	MTG2_YEAST
42	89.5	5.3	462	1	TRPE_THETH
43	88.5	5.2	1103	1	KEIC_HUMAN
44	88	5.2	513	1	NOR1_ALCEB
45	88	5.2	525	1	NAB2_MOUSE

#### ALIGNMENTS

RESULT 1  
MUTY\_SALTY STANDARD; PRT: 350 AA.  
ID ID  
AC 005869;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE A/G-specific adenine glycosylase (EC 3.2.2.-).  
GN MUTY OR MUTB OR STM3110.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_Taxid=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GM1803;  
RA MEDLINE=93123173; PubMed=8419300;  
RA Desiraju V., Shanabroch W.G., Lu A.L.;  
RT "Nucleotide sequence of the Salmonella typhimurium mutb gene, the  
RT homolog of Escherichia coli mutY.";  
RL J. Bacteriol. 175:541-543(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SCS1412 / ATCC 700720;  
RA MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulyanay E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2.";  
RL Nature 413:852-856(2001).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: M86634; AAA27165.1; -  
CC EMBL: AE008843; AAL21985.1; -  
CC PIR: A40647; A40647.  
CC HSSP: P17802; 1MUN.  
CC ScyGene: SG10238; muty.  
CC InterPro: IPR003265; Endo\_3c.  
CC InterPro: IPR004035; EndoIII\_FCL.  
CC InterPro: IPR004036; EndoIII\_Hnh.  
CC InterPro: IPR003651; Fes\_bnd.

```

DR InterPro: IPR000445; Hhh.
DR InterPro: IPR003583; Hhh.
DR InterPro: IPR005760; MutY.
DR Pfam: PF00730; Hhh-GPD; 1.
DR SMART: SM00633; HHH; 1.
DR SMART: SM00478; ENDO3C; 1.
DR SMART: SM00525; FES; 1.
DR SMART: SM00278; HHH1; 1.
DR TIGRfam: TIGR01084; mutY; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; 1.
DR PROSITE: PS01155; ENDONUCLEASE_III_2; 1.
DR DNA repair: Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S;
KW Complete proteome.
FT METAL 192
FT METAL 199
FT METAL 202
FT METAL 208
SQ SEQUENCE 350 AA; 39409 MW; 5C55DB4D7B7BB69F CMC64;

Query Match
Best Local Similarity 28.5%; Score 484.5; DB 1; Length 350;
Matches 126; Conservative 58; Mismatches 115; Indels 49; Gaps 14;

OY 8 LAAVYRENAR-PLPRGKDPYRVLSSEVLLOOTRVQALPYRRFLREPTLKALAAS
DB 10 VLDWYDKYRKLTLPQWIKTKPTKWLSEVLMLOOTQTVTYPFERPMARFPVVDLANAP
OY 67 LEEVRYWQAGYTRRAEHLRLARSYE-----ELPPSAELRGPGCPYTAANAVSIA
DB 70 LDEVLMWTGLGYTAARNRLHKAQOAVTLHGGEFPQFAETALALPGVGRSTAGALISLA
OY 122 FGEVAAVDSNVRRVLSRLFARES--PKEKE-----LPAAGLLP-EGVDPGVWMOALMEL
DB 130 LGRKHYPLTDGNKRVILARCYAVSGMPGKEVENTLMTLSEQVTPARVER--ENQAMMDL
OY 176 GATVCLPKPRGACGL--GACRCKEAPGRYPAPRRKRAKEEELVALYLGGKGVHLER
DB 188 GAAYCTRSKPKCTLCPLONGCJAANAHESRYRPEKPKQTLPRFTGYELL--OHQOE
OY 234 L-----EGRQGLGVPLFPPE-ELPGREAAQVNSRPLGAV--RHATLHRLRY--
DB 244 IFLAQRPSPGIMGGLYCFPPAFARDELREMLAQRIHNAADNLQOLNAFRHTFSHFLDLYP
OY 281 -----EVRGALMEGEGEDP---WKRPLPKIMKRVLRKALPL
DB 304 MMLPVSSILACMDEGSALMTNLNLAQPPSVGLAAVERLLOQ-LRTGAPV

```

```

RESULT 2
MUTY_ECOLI
AC P17802; STANDARD; PRT; 350 AA.
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE A/G-specific adenine glycosylase (EC 3.2.2.-).
GN MUTY OR MICA OR A2961.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-90326504; PubMed-2197596;
RT "MutY, an adenine glycosylase active on G-A mispairs, has homology to
RT endonuclease III."
RL Nucleic Acids Res. 18:3841-3845(1990).
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-91161503; PubMed-2001994;

```

```

RA Tsai-Wu J.-J., Radiceella J.P., Lu A.-L.;
RT "Nucleotide sequence of the Escherichia coli mica gene required for
RT A/G-specific mismatch repair: identity of mica and mutY."
RL J. Bacteriol. 173:1902-1910(1991).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN-K12; MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blatterner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick N.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN (4)
RP X-RAY CRYSTALLOGRAPHY (1.2 ANGSTROMS).
RX MEDLINE-99061333; PubMed-9846876;
RA Guan Y., Manuel R.C., Atwal A.S., Parikh S.S., Mol C.D., Miller J.H.,
RA Lloyd S., Tainer J.A.;
RT "MutY catalytic core, mutant and bound adenine structures define
RT specificity for DNA repair enzyme superfamily."
RL Nat. Struct. Biol. 5:1058-1064(1998).
CC -1- FUNCTION: ADENINE GLYCOSYLASE ACTIVE ON G-A MISPAIRS. MUTY ALSO
CC TO THE OXIDATIVELY DAMAGED FORM OF GUANINE: 7,8-DIHYDRO-8-
CC OXOGUANINE.
CC -1- COPACITOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE
CC CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER
CC POSITIONING OF THE ENZYME ALONG THE DNA STRAND.
CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC use by non-profit institutions. There are no restrictions on its
CC modified and this statement as long as its content is in no way
CC entitles requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: X52391; CA36624.1; -
DR EMBL: M59471; AAA72957.1; -
DR EMBL: U28377; AAA69128.1; -
DR PIR: B38535; B38535.
DR PDB: 1MUD; 29-OCT-99.
DR PDB: 1MUN; 26-AUG-99.
DR PDB: 1MUY; 20-AUG-99.
DR PDB: 1KG2; 26-NOV-02.
DR PDB: 1KG3; 26-NOV-02.
DR PDB: 1KG4; 26-NOV-02.
DR PDB: 1KG5; 26-NOV-02.
DR PDB: 1KG6; 26-NOV-02.
DR PDB: 1KG7; 26-NOV-02.
DR PDB: 1KQJ; 10-APR-02.
DR EcoGene; EGI0627; mutY.
DR InterPro: IPR003365; Endo_3C.
DR InterPro: IPR004035; EndoIII_FCL.
DR InterPro: IPR004036; EndoIII_Hhh.
DR InterPro: IPR003651; FES_bind.
DR InterPro: IPR000445; FES_bind.
DR InterPro: IPR003583; HHH.
DR Pfam: PF00730; Hhh-GPD; 1.
DR SMART: SM00633; HHH; 1.
DR SMART: SM00478; ENDO3C; 1.
DR SMART: SM00525; FES; 1.
DR TIGRfam: TIGR01084; mutY; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; 1.
DR PROSITE: PS01155; ENDONUCLEASE_III_2; 1.
KW DNA repair; Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S; 3D-structure;
FT METAL 192
FT METAL 199

```







OS Buchnera aphidicola (subsp. Schizaphis graminum).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Buchnera.

OX NCBI\_TaxID=98794;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=22084549; Pubmed=12089438;

RA Tamas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,

RA Wernegreen J.F., Sandstrom J.P., Moran N.A., Andersson S.G.E.;

RT "50 million years of genomic stasis in endosymbiotic bacteria.";

RL Science 296:2376-2379(2002).

CC -1- FUNCTION: Adenine glycosylase active on G-A and C-A mispairs (By similarity).

CC -1- COFACTOR: Binds a 4Fe-4S cluster which is not important for the catalytic activity, but which is probably involved in the proper positioning of the enzyme along the DNA strand (By similarity).

CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: AE014127; AAM68075.1; -.

DR InterPro: IPR003265; Endo\_3c.

DR InterPro: IPR004035; EndoIII\_FCL.

DR InterPro: IPR004036; EndoIII\_HhH.

DR InterPro: IPR003651; Fes\_bind.

DR InterPro: IPR005760; MutY.

DR Pfam: PF00730; HhH-GPD; 1.

DR Pfam: PF00633; HhH; 1.

DR SMART: SM00478; ENDO3c; 1.

DR SMART: SM00525; FES; 1.

DR TIGRFSMS: TIGR01084; muty; 1.

DR PROSITE: PS00764; ENDONUCLEASE\_III\_1; 1.

DR PROSITE: PS01155; ENDONUCLEASE\_III\_2; FALSE\_NEG.

DR DNA repair; Hydrolyase; Glycosidase; Iron-sulfur; 4Fe-4S;

KW Complete proteome.

FT METAL 192 IRON-SULFUR (4FE-4S) (BY SIMILARITY).

FT METAL 199 IRON-SULFUR (4FE-4S) (BY SIMILARITY).

FT METAL 202 IRON-SULFUR (4FE-4S) (BY SIMILARITY).

FT METAL 208 IRON-SULFUR (4FE-4S) (BY SIMILARITY).

SO SEQUENCE 347 AA; 41301 MW; D6E6279175293587 CRC64;

Query Match 23.6%; Score 402; DB 1; Length 347;

Best Local Similarity 34.2%; Pred. No. 2, 7e-24;

Matches 77; Conservative 59; Mismatches 75; Indels 14; Gaps 5;

QY 1 VEAARKALAWYRENRAP-LPMRGEKDPYRVLYSEVLLQOTRVQALPYRRFLERPTL 59

DB 3 IYVFSQIILNMYHINCRKMLPMKKDKTLKVMISEIMLQOTVKTALPYRKFNISRPNI 62

QY 60 KALAAASLEEVLRVWVGAGYVRRRAEHLRLARSV-----ELRPSFAELGSLGIGYTA 114

DB 63 QSIINOSKLDLIDLWSGLGTYKRAENITYVTKITKEPFOKPTGSDLLKLGIGSTA 122

QY 115 AAVASIFGERVAALDGNVRRVLSRL-----FARSPKEKELPALAAGLLPEGVDGVMN 169

DB 123 GAILSLSLDYFFILEGNVRRKILMRYGIIIGYTEKKIEKMLVLEILIRPH-NGSFSN 181

QY 170 QALMEIGATVCLPKPRPGACPLGAFRCGKEAPG--RYPAARRR 212

DB 182 OGIMDIGALICTPKNPKCNCLPIQKCIAYKEXKMWIKYPLKRRK 226

RESULT 6

ID MUTY\_BUCAT STANDARD; PRT; 350 AA.

AC P57617;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE A/G-specific adenine glycosylase (EC 3.2.2.-).

GN MUTY OR BU552.

OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum

OS symbiotic bacterium).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Buchnera.

OX NCBI\_TaxID=118099;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Tokyo 1998;

RA MEDLINE=20445173; Pubmed=10993077;

RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;

RT "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. Aps.";

RL Nature 407:81-86(2000).

CC -1- FUNCTION: ADENINE GLYCOSYLASE ACTIVE ON G-A AND C-A MISPAIRS (BY SIMILARITY).

CC -1- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: AP001119; BAB13244.1; -.

DR HSPSP; P17802; 1MUTY.

DR InterPro: IPR003265; Endo\_3c.

DR InterPro: IPR004035; EndoIII\_FCL.

DR InterPro: IPR004036; EndoIII\_HhH.

DR InterPro: IPR003651; Fes\_bind.

DR InterPro: IPR000445; HhH.

DR InterPro: IPR005760; MutY.

DR Pfam: PF00730; HhH-GPD; 1.

DR Pfam: PF00633; HhH; 1.

DR SMART: SM00478; ENDO3c; 1.

DR SMART: SM00525; FES; 1.

DR TIGRFSMS: TIGR01084; muty; 1.

DR PROSITE: PS00764; ENDONUCLEASE\_III\_1; 1.

DR PROSITE: PS01155; ENDONUCLEASE\_III\_2; 1.

DR DNA repair; Hydrolyase; Glycosidase; Iron-sulfur; 4Fe-4S;

KW Complete proteome.

FT METAL 192 IRON-SULFUR (4FE-4S) (BY SIMILARITY).

FT METAL 199 IRON-SULFUR (4FE-4S) (BY SIMILARITY).

FT METAL 202 IRON-SULFUR (4FE-4S) (BY SIMILARITY).

FT METAL 208 IRON-SULFUR (4FE-4S) (BY SIMILARITY).

SO SEQUENCE 350 AA; 41589 MW; F4567EF9B2BD1 CRC64;

Query Match 22.3%; Score 378.5; DB 1; Length 350;

Best Local Similarity 26.6%; Pred. No. 1, 9e-22;

Matches 91; Conservative 86; Mismatches 114; Indels 51; Gaps 12;

QY 8 LLAWYRENRAP-LPMRGEKDPYRVLYSEVLLQOTRVQALPYRRFLERPTLKAALAA 66

DB 10 VLMMYHNGKRDLPWQINKTLTYWISSEIMLQOTTVASALPYRKFKFLNPNKISLWDSK 69

QY 67 LEEVLRVWVGAGYVRRRAEHLRLARSVE-----LPPSFAELGSLGIGYTA 121

DB 70 LDVLYLWSGLGYNNRAKNIKYKSAQILKKRYKGIFPOFSNIQILPGIGSTAGAILSL 129

QY 122 FGERVAADVGNVRRVLSRLA-----RESRKEKELPALAAGLLPEGVDGVMN 176

DB 130 LNEFYPLIDGNVRRILVRYRGISGLKDKKIEKKLWMIESTIPRH-NTGKFNQGMWDIG 188

QY 177 ATVCLPKPRPGACPLGAFRC--RGKEAPGRVYPAARRKRAKEERLVALVLGG-----RK 227

```

DB      169 ASICISIPKCTTOLPLKKECIAIIEKKWKEPYLKNIKKTLPKISWFIITIIHNNFWLKK 248
OY      228 GVHLEERLEGROGYVPLPPEELGGRNACVSRPLG-----EVRHALTH 275
DB      249 NTEGE-----IMKELFCPRKFKKE-----EALIMLEKKININTCENMISFPHKSHITLH 300
OY      276 RR---LVEVRGALMEGEGEDPW---KRP---LPLKMKLY 307
DB      301 INPLLRPLPISEFFEKHKIKMYLKNPHQIHGPRVQKIL 342

RESULT 7
GTRM_METTF
ID      GTRM_METTF      STANDARD:      PRT:      221 AA.
AC      P29588;
DT      01-APR-1993 (Rel. 25, Created)
DT      01-APR-1993 (Rel. 25, Last sequence update)
DE      15-SEP-2003 (Rel. 42, Last annotation update)
OS      Possible G-T mismatches repair enzyme (EC 3.2.2.-) (ORF10).
OG      Plasmid pVY1.
OC      Archaea; Euryarchaeota; Methanobacteri; Methanobacteriales;
OC      Methanobacteriaceae; Methanothermobacter.
(1)     NCBI_TaxID=145262;

RP      SEQUENCE FROM N.A.
RA      STRAIN=DSM 3848 / THF;
RA      MEDLINE=93126090; PubMed=1336177;
RT      Noelling J., van Eeden F.J.M., Eggen R.I.L., de Vos W.M.;
RT      "Modular organization of related archaeal plasmids encoding different
RT      restriction-modification systems in Methanobacterium
RT      thermoformicum.";
RL      Nucleic Acids Res. 20:6501-6507(1992).
-1-     FUNCTION: COULD ACT IN DNA REPAIR. RESTRICTION METHYLASE
CC      M METHI, WHICH IS ENCODED BY THIS PLASMID, GENERATES
CC      5-METHYLMETHYLOSINE WHICH IS, ESPECIALLY UNDER THERMOPHILIC
CC      CONDITIONS, SUBJECT TO DENATINATION RESULTING IN G-T MISMATCHES.
CC      -1- THIS PROTEIN COULD CORRECT THESE MISMATCHES.
CC      COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE
CC      CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER
CC      POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE NH/MUTY FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-stb.ch/announce/
CC      or send an email to license@isb-stb.ch).
CC      EMBL, X68366; CA948433.1; -.
DB      PIR: S30312; S30312.
DB      PDB: 1KEA; 23-JAN-02.
DR      InterPro: IPR003265; Endo_3c.
DR      InterPro: IPR004035; EndoIII_FCL.
DR      InterPro: IPR004036; EndoIII_HNH.
DR      InterPro: IPR003651; Res_D1nd.
DR      InterPro: IPR000445; HNH.
DR      Pfam: PF00730; Hnh-GPD; 1.
DR      Pfam: PF00633; HNH; 1.
DR      SMART: SM00478; Endo3c; 1.
DR      SMART: SM00525; FES; 1.
DR      PROSITE: PS00764; ENDONUCLEASE_III_1; 1.
DR      PROSITE: PS01155; ENDONUCLEASE_III_2; 1.
KW      DNA repair; Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S; Plasmid;
KW      Hypothetical protein; 3d-structure.
FT      METAL 197
FT      METAL 197
FT      METAL 204
FT      METAL 207
FT      METAL 213
FT      SEQUENCE 221 AA; 25400 MW; D085FCEELDDC4B62 CRC64;

```

Query Match	16.18	Score 274.5	DB 1	Length 221
Best Local Similarity	31.7%	Pred. No. 1.5e-14		
Matches	63	Conservative 40	Mismatches 85	Indels 11
				Gaps 4
QY	8	LLAWYRNAPPLPMPGCKDPYRIVSEVILLOOTRFEQALPYRRFLTERFTLLKALMAASL	67	
Db	16	ILTFMNTRDRDPFRHTRDPYILITELLTRTAGHKVIYDKFEVKYKCFEDILTKPK	75	
QY	68	EEVYRWQOGGYTRAEHLRLRLASV-----ELPSFAELDGLDPLGDPYTAANVASIA	121	
Db	76	SEIAKDKEIGLSNRAEQLELAVVINDYGVGRPNRRAIIDPLGVAKYTCGAANVKCLA	135	
QY	122	FGGRVAADVGNVRYVLSRLFARE-----SPKKELFLAAGCLPEGVDPGVMNQAALMELGA	177	
Db	136	FGKAAVADVNFVRINRYFGGSTEINLVNHHKALMELATLVPGGCRD-PRILGIMDPSA	194	
QY	178	TYCLPRRPGCGACPLGATC	196	
Db	195	ITICAPKPKCEKCGMSKLC	213	

## RESULT 8

ID	MJTY_AERHY	STANDARD;	PRT;	99 AA
AC	P46230;			

DT	01-NOV-1995	32,	Created)
DT	01-NOV-1995	32,	Last sec

DE A/G-specific adenine glycosylase (EC 3.2.2.27) (rel. 32, last annotation update)

05 *Aeromonas hydrophila*,  
06 *Bacterioides*,  
07 *Brucella*,  
08 *Campylobacter*,  
09 *Citrobacter*,  
10 *Clostridium*,  
11 *Corynebacterium*,  
12 *Enterobacter*,  
13 *Escherichia coli*,  
14 *Haemophilus*,  
15 *Legionella*,  
16 *Mycobacterium*,  
17 *Pseudomonas*,  
18 *Rickettsia*,  
19 *Salmonella*,  
20 *Shigella*,  
21 *Staphylococcus*,  
22 *Streptococcus*,  
23 *Typhimurium*,  
24 *Yersinia*,  
25 *Zoonosis*,  
26 *Zoonosis*,  
27 *Zoonosis*,  
28 *Zoonosis*,  
29 *Zoonosis*,  
30 *Zoonosis*,  
31 *Zoonosis*,  
32 *Zoonosis*,  
33 *Zoonosis*,  
34 *Zoonosis*,  
35 *Zoonosis*,  
36 *Zoonosis*,  
37 *Zoonosis*,  
38 *Zoonosis*,  
39 *Zoonosis*,  
40 *Zoonosis*,  
41 *Zoonosis*,  
42 *Zoonosis*,  
43 *Zoonosis*,  
44 *Zoonosis*,  
45 *Zoonosis*,  
46 *Zoonosis*,  
47 *Zoonosis*,  
48 *Zoonosis*,  
49 *Zoonosis*,  
50 *Zoonosis*,  
51 *Zoonosis*,  
52 *Zoonosis*,  
53 *Zoonosis*,  
54 *Zoonosis*,  
55 *Zoonosis*,  
56 *Zoonosis*,  
57 *Zoonosis*,  
58 *Zoonosis*,  
59 *Zoonosis*,  
60 *Zoonosis*,  
61 *Zoonosis*,  
62 *Zoonosis*,  
63 *Zoonosis*,  
64 *Zoonosis*,  
65 *Zoonosis*,  
66 *Zoonosis*,  
67 *Zoonosis*,  
68 *Zoonosis*,  
69 *Zoonosis*,  
70 *Zoonosis*,  
71 *Zoonosis*,  
72 *Zoonosis*,  
73 *Zoonosis*,  
74 *Zoonosis*,  
75 *Zoonosis*,  
76 *Zoonosis*,  
77 *Zoonosis*,  
78 *Zoonosis*,  
79 *Zoonosis*,  
80 *Zoonosis*,  
81 *Zoonosis*,  
82 *Zoonosis*,  
83 *Zoonosis*,  
84 *Zoonosis*,  
85 *Zoonosis*,  
86 *Zoonosis*,  
87 *Zoonosis*,  
88 *Zoonosis*,  
89 *Zoonosis*,  
90 *Zoonosis*,  
91 *Zoonosis*,  
92 *Zoonosis*,  
93 *Zoonosis*,  
94 *Zoonosis*,  
95 *Zoonosis*,  
96 *Zoonosis*,  
97 *Zoonosis*,  
98 *Zoonosis*,  
99 *Zoonosis*,  
100 *Zoonosis*,  
101 *Zoonosis*,  
102 *Zoonosis*,  
103 *Zoonosis*,  
104 *Zoonosis*,  
105 *Zoonosis*,  
106 *Zoonosis*,  
107 *Zoonosis*,  
108 *Zoonosis*,  
109 *Zoonosis*,  
110 *Zoonosis*,  
111 *Zoonosis*,  
112 *Zoonosis*,  
113 *Zoonosis*,  
114 *Zoonosis*,  
115 *Zoonosis*,  
116 *Zoonosis*,  
117 *Zoonosis*,  
118 *Zoonosis*,  
119 *Zoonosis*,  
120 *Zoonosis*,  
121 *Zoonosis*,  
122 *Zoonosis*,  
123 *Zoonosis*,  
124 *Zoonosis*,  
125 *Zoonosis*,  
126 *Zoonosis*,  
127 *Zoonosis*,  
128 *Zoonosis*,  
129 *Zoonosis*,  
130 *Zoonosis*,  
131 *Zoonosis*,  
132 *Zoonosis*,  
133 *Zoonosis*,  
134 *Zoonosis*,  
135 *Zoonosis*,  
136 *Zoonosis*,  
137 *Zoonosis*,  
138 *Zoonosis*,  
139 *Zoonosis*,  
140 *Zoonosis*,  
141 *Zoonosis*,  
142 *Zoonosis*,  
143 *Zoonosis*,  
144 *Zoonosis*,  
145 *Zoonosis*,  
146 *Zoonosis*,  
147 *Zoonosis*,  
148 *Zoonosis*,  
149 *Zoonosis*,  
150 *Zoonosis*,  
151 *Zoonosis*,  
152 *Zoonosis*,  
153 *Zoonosis*,  
154 *Zoonosis*,  
155 *Zoonosis*,  
156 *Zoonosis*,  
157 *Zoonosis*,  
158 *Zoonosis*,  
159 *Zoonosis*,  
160 *Zoonosis*,  
161 *Zoonosis*,  
162 *Zoonosis*,  
163 *Zoonosis*,  
164 *Zoonosis*,  
165 *Zoonosis*,  
166 *Zoonosis*,  
167 *Zoonosis*,  
168 *Zoonosis*,  
169 *Zoonosis*,  
170 *Zoonosis*,  
171 *Zoonosis*,  
172 *Zoonosis*,  
173 *Zoonosis*,  
174 *Zoonosis*,  
175 *Zoonosis*,  
176 *Zoonosis*,  
177 *Zoonosis*,  
178 *Zoonosis*,  
179 *Zoonosis*,  
180 *Zoonosis*,  
181 *Zoonosis*,  
182 *Zoonosis*,  
183 *Zoonosis*,  
184 *Zoonosis*,  
185 *Zoonosis*,  
186 *Zoonosis*,  
187 *Zoonosis*,  
188 *Zoonosis*,  
189 *Zoonosis*,  
190 *Zoonosis*,  
191 *Zoonosis*,  
192 *Zoonosis*,  
193 *Zoonosis*,  
194 *Zoonosis*,  
195 *Zoonosis*,  
196 *Zoonosis*,  
197 *Zoonosis*,  
198 *Zoonosis*,  
199 *Zoonosis*,  
200 *Zoonosis*,  
201 *Zoonosis*,  
202 *Zoonosis*,  
203 *Zoonosis*,  
204 *Zoonosis*,  
205 *Zoonosis*,  
206 *Zoonosis*,  
207 *Zoonosis*,  
208 *Zoonosis*,  
209 *Zoonosis*,  
210 *Zoonosis*,  
211 *Zoonosis*,  
212 *Zoonosis*,  
213 *Zoonosis*,  
214 *Zoonosis*,  
215 *Zoonosis*,  
216 *Zoonosis*,  
217 *Zoonosis*,  
218 *Zoonosis*,  
219 *Zoonosis*,  
220 *Zoonosis*,  
221 *Zoonosis*,  
222 *Zoonosis*,  
223 *Zoonosis*,  
224 *Zoonosis*,  
225 *Zoonosis*,  
226 *Zoonosis*,  
227 *Zoonosis*,  
228 *Zoonosis*,  
229 *Zoonosis*,  
230 *Zoonosis*,  
231 *Zoonosis*,  
232 *Zoonosis*,  
233 *Zoonosis*,  
234 *Zoonosis*,  
235 *Zoonosis*,  
236 *Zoonosis*,  
237 *Zoonosis*,  
238 *Zoonosis*,  
239 *Zoonosis*,  
240 *Zoonosis*,  
241 *Zoonosis*,  
242 *Zoonosis*,  
243 *Zoonosis*,  
244 *Zoonosis*,  
245 *Zoonosis*,  
246 *Zoonosis*,  
247 *Zoonosis*,  
248 *Zoonosis*,  
249 *Zoonosis*,  
250 *Zoonosis*,  
251 *Zoonosis*,  
252 *Zoonosis*,  
253 *Zoonosis*,  
254 *Zoonosis*,  
255 *Zoonosis*,  
256 *Zoonosis*,  
257 *Zoonosis*,  
258 *Zoonosis*,  
259 *Zoonosis*,  
260 *Zoonosis*,  
261 *Zoonosis*,  
262 *Zoonosis*,  
263 *Zoonosis*,  
264 *Zoonosis*,  
265 *Zoonosis*,  
266 *Zoonosis*,  
267 *Zoonosis*,  
268 *Zoonosis*,  
269 *Zoonosis*,  
270 *Zoonosis*,  
271 *Zoonosis*,  
272 *Zoonosis*,  
273 *Zoonosis*,  
274 *Zoonosis*,  
275 *Zoonosis*,  
276 *Zoonosis*,  
277 *Zoonosis*,  
278 *Zoonosis*,  
279 *Zoonosis*,  
280 *Zoonosis*,  
281 *Zoonosis*,  
282 *Zoonosis*,  
283 *Zoonosis*,  
284 *Zoonosis*,  
285 *Zoonosis*,  
286 *Zoonosis*,  
287 *Zoonosis*,  
288 *Zoonosis*,  
289 *Zoonosis*,  
290 *Zoonosis*,  
291 *Zoonosis*,  
292 *Zoonosis*,  
293 *Zoonosis*,  
294 *Zoonosis*,  
295 *Zoonosis*,  
296 *Zoonosis*,  
297 *Zoonosis*,  
298 *Zoonosis*,  
299 *Zoonosis*,  
300 *Zoonosis*,  
301 *Zoonosis*

0C Aeromonadaceae; Aeromonas.  
0X NCBI TaxID=644: Gammaproteobacteria; Aeromonadales;

RP	SEQUENCE FROM N
RN	(11)

SRALN=Anb5;  
MEDLINE=95050248; PubMed=7961440.

RT "Isolation and characterization of a second

J. Bacteriol. 176:6819-6826(1994):

(BY SIMILARITY).  
-1- COFACTORS. DIVERSITY

CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER POSITIONING OF THE ENZYME.

----- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY. -----

-----  
This Swiss-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the University of Basel.

European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its contribution to the advancement of the field is acknowledged.

entitles requires a license agreement (See <http://www.ietf.org>)

----- (continued on side 2) -----  
 C  
 -----  
 EMRI: Y81473. VOT  
 B  
 -----

```

R      HSSP; P17802; 1MUN.
R      InterPro: IPR003265. Ex4c 3-

```

InterPro; IPR004035; EndoIII\_FCL.  
InterPro; IPR004036; EndoIII\_Hbu

PROSITE; PS00764; ENDONUCLEASE TTI 1. PABTTAT

DNA repair; Hydrolase; Glycosidase; Iron-sulfur; AEC-AS

SEQUENCE 99 AA; 11692 MW; 0806BDC67E1B0CB6 CRC64.

Query Match	14.78;	Score 250.5;	DB 1;	Length 99;
Best Local Similarity:	14.78;			

Matches	Conservative	Mismatches	Indels	Gaps
45;	19;	22;	1;	1.

ID	ENTRY	STANDARD	PRG	213 AA
0Y	8 LLAAYREAR-PLEWRGCKDDRYLVSVLLQOQTEQALRYRRELEPPLTKLAAS			66
Db	13 ILDMYOLHGRKTLPEMOODKTPYRWVSEIMLQOYAVIVYQRFMARFPVQALQAP			72
0Y	67 LEEVLRVWQAGAGYRYRAEHLRLARSV			93
Db	73 IDEVLHHHTGTGYYARARNLRKKAQOI			99
RESULT 9				
END3_THEME	STANDARD	PRG	213 AA	
AC	09WK0:			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Endonuclease III (EC 4.2.99.18) (DNA-(apurinic or apyrimidinic site) lyase).			
GN	NTX OR TM0366.			
OS	Thermotoga maritima.			
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.			
OX	NCBI_TaxID=23367;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MS8 / DSM 3109;			
RX	MEDLINE=99287316; PubMed=10360571;			
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,			
RA	Halt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,			
RA	McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,			
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,			
RA	Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,			
RA	Saizberg S.L., Smith H.O., Venter J.C., Fraser C.M.;			
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima."			
RL	Nature 399:323-329(1999).			
CC	-1- FUNCTION: HAS BOTH AN APURINIC AND/OR APYRIMIDINIC ENDONUCLEASE ACTIVITY AND A DNA N-GLYCOSYLASE ACTIVITY. INCISES DAMAGED DNA AT CYTOSINES, THYMINES AND GUANINES. ACTS ON A DAMAGED STRAND, 5' FROM THE DAMAGED SITE (BY SIMILARITY).			
CC	-1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or apyrimidinic site in DNA is broken by a beta-elimination reaction, leaving a 3'-terminal unsaturated sugar and a product with a terminal 5'-phosphate.			
CC	-1- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE NTH/MOTY FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce">http://www.isb-sib.ch/announce</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; AE001716; AAD35453.1; -			
DR	PIR; F72387; F72387.			
DR	HSSP; P20625; 2ABK.			
DR	TIGR; TM0366.			
DR	InterPro; IPR003265; Endo_3c.			
DR	InterPro; IPR004035; EndoIII_FCL.			
DR	InterPro; IPR004036; EndoIII_HH.			
DR	InterPro; IPR003651; Res_bind.			
DR	InterPro; IPR000445; HH1.			
DR	InterPro; IPR003583; HH1_1.			
DR	InterPro; IPR005759; Nth.			
DR	Pfam; PF00730; Hnh-GPD; 1.			
DR	Pfam; PF00633; HH1; 1.			
DR	SMART; SMO0478; Endo3c; 1.			
DR	SMART; SMO0525; FES; 1.			
DR	SMART; SMO0278; HH1; 2.			
DR	TIGRfams; TIGR01083; nth; 1.			

```
DR PROSITE PS00764; ENDONUCLEASE_III_1;
KW Hydrolase; Nuclease; Endonuclease; Multifunctional enzyme; DNA repair;
FT METAL 180 180 IRON-SULFUR (4FE-4S) Complete proteome.
FW METAL 187 187 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 190 190 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 196 196 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 213 AA; 24928 MW; D9716CD6EFA8D4 CRC64;
Query Match 13.4%; Score 227; DB 1; Length 213;
Best Local Similarity 34.4%; Pred. No. 7,3e-11;
Matches 66; Conservative 33; Mismatches 71; Indels 22; Gaps 7;;
QY 24 EKDEPYRLVSEVLLOQTREVALPPYRRFLERFPFLKALAASLEEVLRWOGAGCYR-R 82
DB 19 ETDPEFRVLISIVLSQSRTEDENTEKASKLFFVEYRPOLAKAKPEDLDYLKESGMRYHQ 78
OY 83 AEHLHRLARSVE-----LPSPFAELRKPLGLGPTTAAVAASIANGENAAVDGVNRYLV 137
DB 79 AERIIVEISRIVERKYGRVPDSLEDLKLPGRGKTAINIYVMWGKKPALAVDVHVARIS 138
OY 138 SRL--FARESKEKEFLFALAOGLPPEGVDPGW---NOALMELGATVCCLKPRRCGACPL 192
DB 139 NRLGWKTRTPEETE--EALKKLLEP----DLMPGTINSMEVFGRRIKPKNPCLCEECFL 192
OY 193 GAFC-----RGK 199
DB 193 KNHCIFYRRRK 204
RESULT 10
UNVEN_MICLU STANDARD: PRT: 279 AA.
ID UNVEN_MICLU AC PA6303:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ultraviolet N-glycosylase/AP lyase (UV-endonuclease) (Pyrimidine dimer glycosylase).
GN PG5.
OS Micrococcus luteus (Micrococcus lysodeikticus).
OC Bacteria; Actinobacterila; Actinobacteriales; Micrococcales; Micrococcaceae; Micrococcus.
OX NCBI_TaxID=1270;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-35.
RC STRAIN=MTC 4698; PubMed=7559510; MEDLINE=96007490.
RX Pietersen C.E., Prince M.A., Augustine M.L., Dodson M.L., Lloyd R.S.; "Purification and cloning of Micrococcus luteus ultraviolet endonuclease, an N-glycosylase/basic lyase that proceeds via an imino enzyme-DNA intermediate."; J. Biol. Chem. 270:23475-23484(1995).
CC -! FUNCTION: HAS BOTH, AN APURINIC AND/OR APYRIMIDINIC ENDONUCLEASE ACTIVITY AND A DNA N-GLYCOSYLASE ACTIVITY. INITIATES REPAIR AT CIS-SYN PYRIMIDINE DIMERS. PROCEEDS VIA AN IMINO ENZYME-DNA INTERMEDIATE.
CC -! MISCELLANEOUS: READTHROUGH OF THE TERMINATOR UAG OCCURS BETWEEN CODONS FOR GLY-268 AND ALA-270. TWO FORMS OF 31 KDA AND 32 KDA HAVE BEEN DETECTED.
CC -! SIMILARITY: BELONGS TO THE NRH/MUTY FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on lists CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licensel@isb-sib.ch).
CC EMBL; U22181, AAA86508.1; ALT_TERM.
OR HSSP; P20625, ZABK.
```

UR EMBL: 250874; CAA90766.1; -  
DR PIR: T24131; T24131.  
HSSD: p00655

PROSITE; PS00764; ENDONUCLEASE\_III 1: 1

Best Local Similarity 10.48; Score 176; DB 1; length 259;  
Matches 59; Pred. No. 8 8a-07, 32.08;

107 102 107 102  
245 GWIKISIPKTKAL-EILPK---SEMPINHLVLVGFGMQCQPVPRPKCGTC---LC 199

riable endonuclease III (EC 4.2.99.18) (DNA-(apurinic or apyrimidinic site) lyase).

RA Brund C., Sorokin A., Serror P., Ehrlich S.D.;  
 RT "Nucleotide sequence of the Bacillus subtilis dnaD gene."  
 RL Microbiology 141:321-322(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168 / Marburg;  
 RX MEDLINE=96349105; PubMed=8760912;  
 RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,  
 Serror P.;  
 RT "Sequence analysis of the Bacillus subtilis chromosome region between  
 the serA and kgd loci cloned in a yeast artificial chromosome."  
 RL Microbiology 142:2005-2016(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Bertorello M.G., Bessieres P., Bolotin A., Borchert S.,  
 Borst R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 Chol S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
 Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
 Giuseppe G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,  
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 Joris B., Karamata D., Kasahara Y., Kieffer-Blanchard M., Klein C.,  
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Masuda S., Meuel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
 Parro V., Pohl T.M., Portetalle D., Potwolik S., Prescott A.M.,  
 Prescan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Rocha E., Roche R., Rose M., Sadale Y.,  
 Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 Sekiuchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,  
 Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,  
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 Viari A., Wambit R., Wedler E., Wedler H., Weltenegeger T.,  
 Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis."  
 RL Nature 390:249-256(1997).  
 CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or  
 apyrimidinic site in DNA is broken by a beta-elimination reaction,  
 leaving a 3'-terminal unsaturated sugar and a product with a  
 terminal 5'-phosphate.  
 CC -1- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE  
 CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER  
 POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U11289; AAA80005.1; -;  
 DR EMBL: L47709; AAB8457.1; -;  
 DR EMBL: 299115; CAB14150.1; -;  
 DR PIR: I40525; I40525.  
 DR HSSP: P20625; 2ABK.  
 DR Subtilist; BG10956; nth.  
 DR InterPro; IPR003265; Endo\_3c.  
 DR InterPro; IPR004036; EndoIII\_FCL.  
 DR InterPro; IPR004036; EndoIII\_Hhn.

DR InterPro; IPR003651; Fes\_bind.  
 DR InterPro; IPR000445; HHH\_1.  
 DR InterPro; IPR003583; HHH\_1.  
 DR InterPro; IPR005759; Nch.  
 DR Pfam; PF00730; Hhh-GPD; 1.  
 DR Pfam; PF00633; Hhh; 1.  
 DR SMART; SM00478; ENDO3c; 1.  
 DR SMART; SM00525; FES; 1.  
 DR SMART; SM00278; Hhh1; 1.  
 DR TIGRfams; TIGR01083; nth; 1.  
 DR PROSITE; PS00764; ENDONUCLEASE\_III\_1; 1.  
 DR PROSITE; PS01155; ENDONUCLEASE\_III\_2; 1.  
 KW Hydroxylase; Nuclease; Endonuclease; Multifunctional enzyme; DNA repair;  
 KM Glycosidase; Lyase; Iron-sulfur; 4Fe-4S; Complete proteome.  
 FT METAL 189 189 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 196 196 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 199 199 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 205 205 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 SQ SEQUENCE 219 AA; 25000 MW; C38C66DA7948BA40 CRC64;  
 Query Match 10.18; Score 172; DB 1; Length 219;  
 Best Local Similarity 28.18; Pred. No. 1.5e-06;  
 Matches 56; Conservative 30; Mismatches 65; Indels 48; Gaps 6;  
 QY 26 DPYRVLYSEVLQO-----TRVEQALPYRFRLEPTIKALAAALEVLRWQAGAY 79  
 Db 29 NPPELVVAVALASQCDALNRYTKL-----FQKKREDYLAFLLELODKISGL 82  
 QY 80 YR-RAEHLIRLANSV-----EELPPSFALRGILPGIPYAAVAATFGERVAADGVN 133  
 Db 83 YRNKAKNIQIKLTKMIEDYGEVDRDELVKLPVGRKTAANYVSAFVPAIAYDTNV 142  
 QY 134 RYVLSRL-----FARESPKKEKLFALAGLLPESVDPGVNNQALMEIGAT 178  
 Db 143 ERYSKRIGRCGRWDSVLEVEKTLMKRYPK-----DMSVTHHRIIFGRY 187  
 QY 179 VCLPRPGACPGAFRC 197  
 Db 188 HCKAQSPRCACPELISCR 206  
 RESULT 13  
 ID Y613.METJA STANDARD; PRT; 344 AA.  
 AC 058030;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative endonuclease MJ0613.  
 GN MJ0613.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 CC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton A.A., Gocayne J.D.,  
 Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 Scott J.L., Geoghegan N.S.M., Meldrum J.F., Ruhnann J.L., Nguyen D.,  
 Uiterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 Cotton M.D., Roberts R.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii."  
 RL Science 273:1058-1073(1996).  
 CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -



SEQUENCE	344 AA;	40567 MW;	BC3BEFDE778B85 CRC64;
METAL	184	184	IRON-SULFUR (4FE-4S) (BY SIMILARITY)
METAL	130	190	IRON-SULFUR (4FE-4S) (BY SIMILARITY)
METAL	133	193	IRON-SULFUR (4FE-4S) (BY SIMILARITY)
METAL	201	201	IRON-SULFUR (4FE-4S) (BY SIMILARITY)
SEQUENCE	344 AA;	40567 MW;	BC3BEFDE778B85 CRC64;

Query Match	10.1%	Score 171	DB 1	Length 344
Best Local Similarity	28.4%	Prod. No. 3e-06		
Matches	55	Conservative 41	Mismatches 80	Indels 18
				Gaps 7
QY	8	LLAARYRENAPDLPMRGECKDPYRVLSSEVLLQOTRVECALPYRRRLERFPFTLLAAMSL		
Db	7	LLKRLKNNAVETIADKDPKPEVLLSTIISARTKDEVTVEYSGKLPEIKRDVDLLINDE		
QY	68	EEVYLRVMOGGYVR-RAEHLRLARLSVEE-----LPPSFALDGLDGLGYTAAVAASIA		
Db	67	EKLADLIYPGTYKNAKKKLLAIIKENNGKVPDSEBELLLKVRKRLANVITLA		
QY	122	FGGRVAVAGGNRRVRLSR-LFARSPKEKELEFALAGLLPEGVDSGYV---NQLMBELG		
Db	127	FNNDDGCVDTYHARICRNMEITVDTEPEETE-FELIRKKLPKK-----YKVINLLNLLVEG		
QY	177	ATYCLPKRRRCAC		
Db	181	RETCSSK-SKCDKC		

	RESULT 14
ID	NT_H1 HUMAN
AC	P78549; Q99566; Q99794; Q9BPX2; STANDARD; 312 AA.
DT	28-FEB-2003 (Rel. 41, Created)
DT	28-FEB-2003 (Rel. 41, Last sequence update)
DT	15-SEP-2003 (Rel. 42, Last annotation update)
DE	Endonuclease IIR-like protein 1 (EC 4.2.99.18).
GN	ENDONUCLEASE IIR OR NT_H1 OR OC7S.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX	NCBI_TaxID=9606;
RP	SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RA	MEDLINE=9714402; PubMed=8990169;
RA	Asplimwall R., Rothwell D.G., Roldan-Arjona T., Anselmino C.,

RA Ward C.J., Cheadle J.P., Sampson J.R., Lindehl T., Harris P.C.,  
RA Hickson I.D.;  
RT "Cloning and characterization of a functional human homolog of  
RT Escherichia coli endonuclease III."; *Proc. Natl. Acad. Sci. U.S.A.* 94:109-114(1997).  
RL [2]  
RN  
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.  
RC TISSUE=Placenta;  
RX MEDLINE=99051340; PubMed=9831664;  
RA Imai K., Sarker A.H., Akiyama K., Ikeda S., Yao M., Tsutsui K.,  
RA Shiomori T., Seki S.;  
RT "Genomic structure and sequence of a human homologue (NTHLI/NTHI) of  
RT Escherichia coli endonuclease III with those of the adjacent parts of  
RT TSC2 and SLG9A3R2 genes."; *Gene* 222:287-295(1998).  
RL [3]  
RN  
RP SEQUENCE FROM N.A.  
RA Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,  
RA Robinson D., Jones W., Buckingham J., Chasteen L., Thompson S.,  
RA Goodwin L., Bryant J., Tesmer J., Melnick L., Longmire J., White S.,  
RA Ueng S., Tatum O., Campbell C., Fawcett J., Maltbie M., Mistr M.,  
RA Deaven L.;  
RT "Sequencing of human chromosome 16p13.3";  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Rieder M.J., Braun A.C., Montoya M.A., Chung M.-N., Nguyen C.P.,  
RA Nguyen D.A., Livingston R.J., Poel C.L., Robertson P.D.,  
RA Schackwitz W.S., Sherwood J.K., Wiltrak L.A., Nickerson D.A.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP  
RP SEQUENCE OF 6-312 FROM N.A., FUNCTION, ACTIVE SITE, AND MUTAGENESIS OF  
RP LYS-220.  
RC  
RC TISSUE=Bone marrow;  
RX MEDLINE=98370989; PubMed=9705289;  
RA Ikeda S., Bistwas T., Roy R., Izumi T., Boldogh I., Kurosky A.,  
RA Sarker A.H., Seki S., Mitra S.;  
RT "Purification and characterization of human NTHI, a homolog of  
RT Escherichia coli endonuclease III. Direct identification of Lys-212 as  
RT the active nucleophilic residue."; *J. Biol. Chem.* 273:21585-21593(1998).  
RL [6]  
RN  
RP SEQUENCE OF 7-312 FROM N.A., AND FUNCTION.  
RC  
RC TISSUE=Spleen;  
RX MEDLINE=97197855; PubMed=9045706;  
RA Hilbert T.P., Chung W., Boorstein R.J., Cunningham R.P., Teehor G.W.;  
RT "Cloning and expression of the cDNA encoding the human homologue of  
RT J. Biol. Chem. 272:6733-6740(1997).  
RL [7]  
RN  
RP SEQUENCE OF 8-312 FROM N.A.  
RC  
RC TISSUE=Lung;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenman C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buckow K.H., Schaefer C.F., Bhat N.J.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diachenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshlygi S., Carninci P., Prange C.,  
RA Rabb S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mollaby S.J.,  
RA Boesk S.A., McEwan P.J., McKernan K.J., Malek J.A., Guneratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huily S.W.,  
RA Villalon D.K., Mundy D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Keltman K., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences"; *Proc. Natl. Acad. Sci. U.S.A.* 99:16899-16903(2002).



RN [8] SEQUENCE OF 9-312 FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND  
 RP DEVELOPMENTAL STAGE.  
 RX MEDLINE=20342089; PubMed=10882850;  
 RA Luna L., Bjoras M., Hoff E., Rognes T., Seeberg E.;  
 RT "Cell-cycle regulation, intracellular sorting and induced  
 overexpression of the human NTH1 DNA glycosylase involved in removal  
 of formidopyrimidine residues from DNA.";  
 RL Mutat. Res. 460:95-104(2000).  
 RN [9]  
 RP SUBCELLULAR LOCATION.  
 RA MEDLINE=22419222; PubMed=12531031;  
 RX Ikeda S., Kohmoto T., Tabata R., Seki Y.;  
 RT "Differential intracellular localization of the human and mouse  
 endonuclease III homologs and analysis of the sorting signals.";  
 RL DNA Repair 1:847-854(2002).  
 CC -1- FUNCTION: Has both an apurinic and/or apyrimidinic endonuclease  
 activity and a DNA N-glycosylase activity. Incises damaged DNA at  
 cytosines, thymines and guanines. Acts on a damaged strand, 5'  
 from the damaged site. Required for the repair of both oxidative  
 DNA damage and spontaneous mutagenic lesions.  
 CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or  
 apyrimidinic site in DNA is broken by a beta-elimination reaction,  
 leaving a 3'-terminal unsaturated sugar and a product with a  
 terminal 5'-phosphate.  
 CC -1- COFACTOR: Binds a 4Fe-4S cluster which is not important for the  
 catalytic activity, but which is probably involved in the proper  
 positioning of the enzyme along the DNA strand.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Widely expressed with highest levels in heart  
 and lowest levels in lung and liver.  
 CC -1- DEVELOPMENTAL STAGE: Expression levels are regulated during the  
 cell cycle with increased levels during early and mid S-  
 phase.  
 CC -1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.  
 CC -1- CAUTION: It is uncertain whether Met-1, Met-9 or Met-16 is the  
 initiator.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U79718; AAB41534.1; -;  
 DR EMBL: AB014460; BAA32695.1; ALT\_INT.  
 DR EMBL: AC005600; AAC34209.1; -;  
 DR EMBL: AF498098; AAM11786.1; -;  
 DR EMBL: AB001575; BAA19413.1; ALT\_INT.  
 DR EMBL: U81285; AAC51136.1; ALT\_INT.  
 DR EMBL: BC003014; AAH03014.1; ALT\_INT.  
 DR EMBL: BC000391; AAH00391.1; ALT\_INT.  
 DR EMBL: Y09687; CAA70865.1; -;  
 DR GeneID: HGNC:8028; NTHL1.  
 DR MIM: 602656; -;  
 DR GO: GO:0005634; C:nucleus; IDA.  
 DR GO: GO:0019104; F:DNA N-glycosylase activity; IDA.  
 DR GO: GO:0004519; F:endonuclease activity; TAS.  
 DR GO: GO:0006284; P:base-excision repair; TAS.  
 DR GO: GO:0006296; P:nucleotide-excision repair, DNA incision, 5...; IDA.  
 DR InterPro: IPR003265; Endo3c.  
 DR InterPro: IPR004036; EndoIII\_Hhm.  
 DR InterPro: IPR003651; Fes\_Bldn.  
 DR Pfam: PF00730; Hhm-GPD; 1.  
 DR Pfam: PF00633; Hhm; 1.  
 DR SMART: SM00478; ENDO3c; 1.  
 DR SMART: SM00525; FES; 1.  
 DR PROSITE: PS00764; ENDONUCLEASE\_III\_1; FALSE\_NEG.  
 DR PROSITE: PS01155; ENDONUCLEASE\_III\_2; 1.  
 KW Hydrolase; Nuclease; Endonuclease; Multifunctional enzyme; DNA repair;

KW Glycosidase; Lyase; Iron-sulfur; 4Fe-4S; Nuclear protein.  
 FT ACT\_SITE 220 220 NUCLEOPHILE IN THE N-GLYCOSYLASE  
 FT REACTION.  
 FT METAL 290 290 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 297 297 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 300 300 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 306 306 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT MUTAGEN 220 220 K->R: 85-FOLD REDUCTION IN ACTIVITY.  
 FT MUTAGEN 220 220  
 FT CONFLICT 9 10 MT -> TS (IN REF. 8).  
 FT CONFLICT 78 78 MISSING (IN REF. 8).  
 FT CONFLICT 151 151 M -> I (IN REF. 1).  
 FT CONFLICT 160 160 T -> A (IN REF. 1).  
 SQ SEQUENCE 312 AA; 34389 MW; 379816A1E0B4505050 CRC64;  
 Query Match 9.7%; Score 164.5; DB 1; Length 312;  
 Best Local Similarity 28.1%; Pred. No. 8.66-06;  
 Matches 56; Conservative 38; Mismatches 80; Indels 25; Gaps 7;  
 QY 12 YRENARPLPWRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFTLKALAAASLEVL 71  
 Db 119 YDSSAPFKVR-----YQVLISLMSQTKDQVAGAMQRLRAGLVDSITQDDMTLG 173  
 QY 72 RVWGAGCYRR-RAEHLRLARSYE-----ELPSFAELRGIPGIPYTAASVIAEGE- 124  
 Db 174 KLIYVGFMRSKYKVIKQTSALIQHYGDIPIASVAVELVALPGVKMAHLAMVAVMGTV 233  
 QY 125 RVAADVQNRVRL-----FARSPKEKELFALAGLLEPGVPGW---NQLMELGA 177  
 Db 234 SGIAVDVHVRHRIARLMTKATKSPBE-----TRAALDEWLPRELMHEINGLLVGFQ 287  
 QY 178 TVCLPKRRCACPLGAF 196  
 Db 288 QTCLPVPRCHACLNQALC 306  
 RESULT 15  
 NTH1\_MOUSE  
 ID NTH1\_MOUSE STANDARD; PRT; 300 AA.  
 AC 035980;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Endonuclease III-like protein 1 (EC 4.2.99.18).  
 GN NTHL1 OR NTHL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RC STRAIN=BALE/C;  
 RX MEDLINE=98417643; PubMed=9743625;  
 RA Sarker A.H., Ikeda S., Nakano H., Terato H., Ide H., Imai K.,  
 RA Akiyama K., Tsutsui K., Bo Z., Kubo K., Yamamoto K., Yasui A.,  
 RA Yoshida M.C., Seki S.;  
 RT "Cloning and characterization of a mouse homologue (mNthl1) of  
 Escherichia coli endonuclease III.";  
 RL J. Mol. Biol. 282:761-774(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Luna L., Bjoras M., Rognes T., Hoff E., Seeberg E.;  
 RT "Complete genomic DNA sequence of the Mus musculus endonuclease III  
 homologue I gene (NTH1)."  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cecum;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nakai T., Osato N., Saito R., Suzuki H., Yamana K.I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Buit C., Hume D.A., Quackenbush J.;

RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chochis C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Kongaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillal R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shindala K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita K.,  
RA Viterbo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilmink L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zengelen M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakakume N., Seto K.,  
RA Shiraki T., Waki K., Kawai J., Alzawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa A.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RT Birney E., Hayashizaki Y.,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RN Mature 420:563-573(2002).  
(4)  
RP SUBCELLULAR LOCATION.  
RX MEDLINE=22419222; PubMed=12531031;  
RT Ikeda S., Kohmoto T., Tabata R., Seki Y.;  
RT "Differential intracellular localization of the human and mouse  
RT endonuclease III homologs and analysis of the sorting signals.";  
CC -!- FUNCTION: Has both an apurinic and/or apyrimidinic endonuclease  
CC activity and a DNA N-glycosylase activity. Incises damaged DNA at  
CC cytosines, thymines and guanines. Acts on a damaged strand, 5'  
CC from the damaged site. Required for the repair of both oxidative  
CC DNA damage and spontaneous mutagenic lesions.  
CC -!- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or  
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,  
CC leaving a 3'-terminal unsaturated sugar and a product with a  
CC terminal 5'-phosphate.  
CC -!- COFACTOR: Binds a 4Fe-4S cluster which is not important for the  
CC catalytic activity, but which is probably involved in the proper  
CC positioning of the enzyme along the DNA strand (By  
CC similarity).  
CC -!- SUBCELLULAR LOCATION: Mitochondrial.  
CC -!- TISSUE SPECIFICITY: Widely expressed.  
CC -!- SIMILARITY: BELONGS TO THE NTH/MYH FAMILY.  
CC -!- CAUTION: Ref.3 sequence differs from that shown due to a  
CC frameshift in position 33.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC use by non-profit institutions. There are no restrictions on its  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AB006812; BAA22080.1; -  
DR EMBL: AB000937; BAA28846.1; -  
DR EMBL: AJ001617; CAB5239.1; -  
DR EMBL: Y09688; CAI70866.1; -  
DR EMBL: AK033701; BAC28435.1; ALT\_FRAME.  
DR MGI: MGI:1313275; NCH11.  
DR GO: GO:0005739; C:mitochondrion; IDA.  
DR GO: GO:0005634; C:nucleus; IDA.  
DR GO: GO:0019104; F:DNA N-glycosylase activity; IDA.  
DR GO: GO:0006296; P:nucleotide-excision repair, DNA incision, 5. . .; IDA.  
DR InterPro: IPR003265; Endo3c.  
DR InterPro: IPR004036; EndoIII\_Hhm.  
DR InterPro: IPR003651; FES\_bind.  
DR InterPro: IPR004445; Hhm.

DR Pfam: PF00730; Hhm-GPD; 1.  
DR Pfam: PF00633; Hhm; 1.  
DR SMART: SM00478; ENDO3c; 1.  
DR SMART: SM00525; FES; 1.  
DR PROSITE: PS00764; ENDOUCLEASE\_III\_1; FALSE\_NEG.  
DR PROSITE: PS01155; ENDOUCLEASE\_III\_2; 1.  
KW Hydroxylase; Nuclease; Endonuclease; Multifunctional enzyme; DNA repair;  
KW Glycosylase; Lyase; Iron-sulfur; 4Fe-4S; Mitochondrion  
FT ACT\_SITE 208  
FT METAL 278  
FT METAL 285  
FT METAL 288  
FT METAL 294  
FT CONFLICT 112  
SQ SEQUENCE 300 AA; 33646 MW; A7D771FFD7BC1939 CRC64;  
Query Match 9.5%; Score 161.5; DB 1; Length 300;  
Best Local Similarity 25.8%; Pred. No. 1.4e-05;  
Matches 56; Conservative 41; Mismatches 87; Indels 33; Gaps 6;  
QY 11 WTREMARPLPRGKDP-----YRVIVSEVLLQOTRVEGALPY 49  
Db 80 WOODLANIRIMRSKRDAPVDQGAENHCYDASAPPKVRVQVLSLMSQTKDOVTAGAM 139  
QY 50 RRELEPPTLKALAAASLEFVLRWMOGAGYR-RAEHLRLARSE-----ELPPFAEL 103  
Db 140 QRLARGLFVESILQTDQDTGLRLTYVGFMRKVKYIKQTALIQORVEGDIPASVAEL 199  
QY 104 RGLPGGLPYRAAASVARG-RVAAVDGNVRVLSRLFARESPEKEKELFALAGLPEG 162  
Db 200 VALPGVPPKRAHLAAVAVAGTISGLIADVTHVRIANRL-RMTKMKTKTPETPKNLEEW 257  
QY 163 VDPGVW---NQALMELGATVCLPKRPRGACPLGAF 196  
Db 258 LPRVLWSEVNGLLVGFQGQICLPVPRQACLNKALC 294

Search completed: September 15, 2003, 13:51:14  
job time : 25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 13:49:19 : Search time 30 Seconds  
(without alignments)  
458.368 Million cell updates/sec

Title: US-09-938-901-2

Perfect score: 1700  
Sequence: 1 VEAWRKALLAWRENAKPLP.....VLRKALPLAHAGVPLPDA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/laa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/laa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/laa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/laa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/laa/PTGUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/laa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	505	29.7	403	4	US-09-107-532A-7308
2	484.5	28.5	535	3	US-08-813-574-2
3	477.5	28.1	470	4	US-09-252-991A-2310
4	477	28.1	375	4	US-09-198-452A-421
5	471.5	27.7	350	3	US-08-813-574-9
6	471.5	27.7	350	4	US-09-651-656-37
7	471.5	27.7	350	4	US-09-650-855-37
8	443	26.1	360	4	US-09-134-001C-4684
9	442	26.0	346	4	US-09-328-352-4982
10	274.5	16.1	221	1	US-08-663-023-17
11	274.5	16.1	221	4	US-09-402-859A-10
12	188	11.1	227	4	US-09-107-532A-3843
13	176	10.4	207	2	US-08-808-550-36
14	176	10.4	259	2	US-08-808-550-39
15	172	10.1	209	2	US-08-808-550-31
16	171	10.1	204	2	US-08-808-550-32
17	164.5	9.7	211	2	US-08-808-550-37
18	164.5	9.7	237	2	US-08-808-550-42
19	164.5	9.7	304	2	US-08-808-550-2
20	160	9.4	224	4	US-09-134-001C-3225
21	154	9.1	213	2	US-08-808-550-35
22	145	8.5	207	2	US-08-808-550-29
23	145	8.5	211	2	US-08-808-550-38
24	145	8.5	211	4	US-09-651-656-25
25	145	8.5	211	4	US-09-650-855-25
26	134	7.9	207	2	US-08-808-550-30
27	128	7.5	628	4	US-09-252-991A-30904

28	127.5	7.5	212	4	US-09-252-991A-17572
29	121	7.1	4472	2	US-08-804-227C-2
30	118.5	7.0	776	4	US-09-252-991A-28446
31	117	6.9	230	4	US-09-328-352-5770
32	115.5	6.8	494	4	US-09-252-991A-17295
33	115	6.8	542	4	US-09-252-991A-21793
34	115	6.8	702	4	US-09-252-991A-31609
35	114.5	6.7	425	4	US-09-252-991A-26336
36	113	6.6	316	4	US-09-252-991A-30325
37	111	6.5	675	4	US-09-252-991A-29650
38	111	6.5	1729	4	US-09-553-690-2
39	109.5	6.4	1580	2	US-08-804-227C-11
40	109.5	6.4	1580	2	US-08-804-198-5
41	108.5	6.4	460	4	US-09-328-352-4274
42	107.5	6.3	427	4	US-09-252-991A-30434
43	106	6.2	435	4	US-09-252-991A-24914
44	106	6.2	798	4	US-09-252-991A-25309
45	105.5	6.2	657	4	US-09-252-991A-28001

#### ALIGNMENTS

RESULT 1  
US-09-107-532A-7308  
Sequence 7308, Application US/09107532A  
Patent No. 6583275

GENERAL INFORMATION:  
APPLICANT: Lynn A Doucelte-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FACIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS: ENTEROCOCCUS CORPORATION  
ADDRESS: 100 Beaver Street  
City: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107, 532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085, 598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:  
NAME: Arinello, Pamela Deneke  
REGISTRATION NUMBER: 40, 489  
REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-8277  
TELEFAX: (781)893-5007

INFORMATION FOR SEQ ID NO: 7308:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 403 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...403  
SEQUENCE DESCRIPTION: SEQ ID NO: 7308:  
US-09-107-532A-7308

Sequence 17572, A  
Sequence 2, Appl1  
Sequence 28446, A  
Sequence 5770, Ap  
Sequence 21795, A  
Sequence 31609, A  
Sequence 26336, A  
Sequence 30325, A  
Sequence 29650, A  
Sequence 2, Appl1  
Sequence 11, Appl1  
Sequence 5, Appl1  
Sequence 30434, Ap  
Sequence 30434, A  
Sequence 24914, A  
Sequence 25309, A  
Sequence 28001, A

Query Match 29.7%: Score 505; DB 4; Length 403;  
 Best Local Similarity 32.7%: Pred. No. 1,4e-42;  
 Matches 118; Conservative 66; Mismatches 113; Indels 64; Gaps 11;

```

OY 9 LAMVRENAPRLPMRGKEDPYRVLVSEVLLQOTRVEQALPYRRRLERFPFLKLAASLE 68
DB 32 IGVYDEKRNLFMRNDRDPYRWIMSEIMLQOTRVDYIDYFEMEFPIEELANAPPE 91
OY 69 EVLRWQAGAGYVRAEHLRLARSV-----ELPPSAELRGLPGDGYTAAVAASIAFG 123
DB 92 KLLKAMEGLGYSSRAARNIOAAKQIMSEFDGEMQPEETISSLKGIGPYTGALISIAFG 151
OY 124 ERVAADVGNVRVLSRLPARES-----PKKEELPALAQGLLPGCVDPGVNQAAMLELGTAV 179
DB 132 LPEPAVDGNVRVLSRLPARES-----PKKEELPALAQGLLPGCVDPGVNQAAMLELGTAV 179
OY 180 CLKPRPGACPLGAFG-----RGKAPGRYPA-PRKRAKKEERLVALVLLGRKGVH---- 230
DB 212 CTPLSPKCTCPLOAFCLANKRGIOF--SFYKTKAKPKOVYISALQNHSGAYYFEE 269
OY 231 -----LEREGFQGLGVPLF-----PPEELP-GREAFGV 261
DB 270 RDSQKLLANMTPPMMEVYQOEYERLKKEMETKEPEIDLVDLVAEEQNLPEKQELFW 329
OY 262 RSRPGEVHALTHRLRYEV--RGALMEGEGED-----PKMRP-----LPKIMEKV 306
DB 330 QTRHAGEVTHIFSHLKMVLLFYGRATEGAQDETENKYSKMLPDAPFVSVPPKVMKL 389
OY 307 L 307
DB 390 V 390

```

RESULT 2  
 US-08-813-574-2  
 ; Sequence 2, Application US/06813574  
 ; Patent No. 6013473  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wei, Yung-Fei  
 ; TITLE OF INVENTION: Human Muty  
 ; NUMBER OF SEQUENCES: 18  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SmithKline Beecham Corporation  
 ; STREET: 709 Swedeland Road  
 ; CITY: King of Prussia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19406  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/813,574  
 ; FILING DATE:  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/013,132  
 ; FILING DATE: 11-MAR-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Han, William T  
 ; REGISTRATION NUMBER: 34,344  
 ; REFERENCE/DOCKET NUMBER: ATG50002  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 610-270-5219  
 ; TELEFAX: 610-270-4026  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 535 amino acids  
 ; TYPE: amino acid

STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-813-574-2

Query Match 28.5%: Score 484.5; DB 3; Length 535;  
 Best Local Similarity 36.8%: Pred. No. 2.5e-40;  
 Matches 136; Conservative 46; Mismatches 107; Indels 83; Gaps 13;

```

OY 1 YEAMKALLAWY-RENAPRLPMRG-----ENDPYRVLVSEVLLQOTRVEQALPYRRF 52
DB 79 VTAERGSLDSMYDEKEDLPMRRARDEMDLDRVAVWSEVLMQOTQVATVINYTGW 138
OY 53 LERFPTKALAAASLEEVLVWQAGYVRAEHLRLARS-VEEL---PSPAFEL--RGL 106
DB 139 MQKWPTLQDLASLSLEEVNQMLAGLYSRGRRLQEGARKVVEELGGMRTETLQQL 198
OY 107 PGCLPTAAAVASIAFERVAADVGNVRVLSRLFA-----RESPKEELFALAQGLPE 161
DB 199 PGVGRYTAGASIAFGATGVGVGNVAVRLCRVRAIGADPSSTLVSQQLMGIAQOLV-D 257
OY 162 GVDPGVWQNALMELGATVCLPKPRPGACPLGAFGRKE----- 200
DB 258 PARPGDFNQAMELGATVCTPQRPICQCPVESICRAKQHVQEQLLAGSLSSGSPDVEE 317
OY 201 -AP-----GRYPARRRRRAK-----EERLVALV-----LGRKGVHLE 232
DB 318 CAPNTGQCHLCLPSEPMDQTLGVNFPKRASKPRESSATCYLEPGALGAQILLVQ 377
OY 233 R-LEGRFGILYGP--LFPPEELPGRPAF-----GVNSRPLGEVHALTHRL 278
DB 378 RNSGLLAGLMEFSPVWSPBESQLRKALLQELQWAGPLPATHLRLHGEVYHFSHKL 437
OY 279 RVEVREGALMEGE 290
DB 438 TYGVYGLALEGQ 449

```

RESULT 3  
 US-09-252-991A-23310  
 ; Sequence 23310, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR APPLICATION NUMBER: 1999-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 23310  
 ; LENGTH: 470  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-23310

Query Match 28.1%: Score 477.5; DB 4; Length 470;  
 Best Local Similarity 36.2%: Pred. No. 1.1e-39;  
 Matches 126; Conservative 58; Mismatches 119; Indels 45; Gaps 14;

```

OY 2 EAMRKALLAWY-RENAPRLPMRGKEDPYRVLVSEVLLQOTRVEQALPYRRRLERFPFLK 60
DB 119 EGFNCAVLDMDTRGRDLPMDGIGTPYRWVSEIMLQOTQVSTYLGTFDFRMAALPDVE 178
OY 61 ALAASLEEVLRWQAGYVRAEHLRLA-----RSVEELPSPFAELRGLPGDGYTA 115
DB 179 ALAAADEVLEHLMTGLGYSSRAARNLHKTAQIVERRHAGEPRPDVEQLAEPLGIGSTAG 238
OY 116 AVASIAFERVAADVGNVRVLSRLPARE---SPK-EKELFALAQGLLPGVDPPGVNQ 170

```

Db 239 AIALSLSGTLAPILDDGVKRVRLARYLAQDGVPGEPKVARALMEAEFTPH-ARVNHVYQ 297  
Qy 171 ALMELGATVCLPRPRGACPLGAFCR---GKEAPERYPAPKRRRA-KEERLVALVLLG 225  
Db 298 AMMDLGTATLCTRSKPSCLLPVSGCRHLLGREAD--YPOKPRKALPKQKRLMPLIAN 355  
Qy 226 RKG---VHLERLGFRGGLGVPLFPPEELPGREAAFGVRSRPLGEVR-----HALTHR 276  
Db 356 RDAAILLYRRPSSGLMGLMSLPDL--DDLDGLEPLAARSLALGERRELSGLTHRTSHF 413  
Qy 277 RLVE-----VRGA-----LWEGEGEDPMKRPPLKLMKYLKRA 310  
Db 414 QLAIEPMLVAVGACAPRAVAGCDMLWYLAAPP--RLGLAAPVKLLKRA 460

RESULT 4  
US-09-198-452A-421  
; Sequence 421, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Griffiths, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 421  
; LENGTH: 375  
; TYPE: PRF  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-421

Query Match 28.1%; Score 477; DB 4; Length 375;  
Best Local Similarity 38.6%; Pred. No. 8,6e-40;  
Matches 117; Conservative 44; Mismatches 112; Indels 30; Gaps 7;

Qy 6 KALLARYRENARLPMBGEKDPYRLVSEVLLQOTRVEQALPYRRPLEPPTLKALAA 65  
Db 22 EALKKWEKKNKRSIPNDNPTYSVWSEVMLQOTRAEVVIDYFNOMMEREPPTIESLAA 81  
Qy 66 SLEEVLRVMQAGAGYRRARHRLARSVEE---LPSFAELRGJLGLGPTYAAVASI 120  
Db 82 KEDVDIKLWEGLEYSRAHLLLEGARVMEEPHGKIPDDAISLAQIRGVGPVTHALAF 141  
Qy 121 AFGERVAADVGNVRYLSRLFARESPEKE---LFALAQGLPEGVDPGVNQAALMEL 175  
Db 142 AFKRRRAADVGNVRLVLSRFLFLETSLDLESTRTWVSRIQAALLPH-KSPREYLAELIEL 200  
Qy 176 GATVCLPRPRGACPLGAFCRGKEAPRGYRPAKRRRAE---ERLVALVL-GRKGVH 230  
Db 201 GATIC-KKYVQCHRCVPRQACGAMREKQFVLPVRHARKKVIPLHRLVALVLDGSLVE 259  
Qy 231 LELLEGFRGGLGVPLFPPEELPGREAAFGVRSR-----PLGEVNHALLTHR 277  
Db 260 KRRPKEMAGLYEPPTYEVEPEEGLODIEGFTKKMELSLPFLGINKLEQRAAFTHNK 319  
Qy 278 LRV 280  
Db 320 VHL 322

RESULT 5  
US-08-813-574-9  
; Sequence 9, Application US/08813574  
; Patent No. 6013473  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Ying-Fel  
; TITLE OF INVENTION: Human Muty  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/813,574  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/013,132  
FILING DATE: 11-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Han, William T  
REGISTRATION NUMBER: 34,344  
REFERENCE/DOCKET NUMBER: ATG50002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5219  
TELEFAX: 610-270-4026  
TELEX:  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 350 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-813-574-9

Query Match 27.7%; Score 471.5; DB 3; Length 350;  
Best Local Similarity 35.1%; Pred. No. 2,8e-39;  
Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

Qy 8 LLMARYENAR-PLMBGEKDPYRLVSEVLLQOTRVEQALPYRRPLEPPTLKALAA 66  
Db 10 VLMDYDYGKRTLPWQDKTPYKWLSEVMLQOTQVATVPIPERFMAKRPPTVLDLANP 69  
Qy 67 LEEVLRVMQAGAGYRRARHRLARSVEE---LPSFAELRGJLGLGPTYAAVASIA 121  
Db 70 LDEVLRHMTIGIYARRARNHKAQVATLHGCKFPTFEVVALPVGSGTAGALLSLS 129  
Qy 122 FGERVAADVGNVRYLSRLFARES-PEKE---LFALAQGLPE-GVDPGVNQAALMEL 175  
Db 130 LCKHFTPLDDGVKRVRLARYLAQDGVPGEPKVARALMEAEFTPH-ARVNHVYQ 297  
Qy 176 GATVCLPRPRGACPL-GAFCRGKEAPRGYRPAKRRRAE---ERLVALVL-GRKGVH 231  
Db 188 GAMICTRSKPKCSLCPLONCIAAANSMALYCKKPKQPLPERTGYFLLDHEDVLLA 247  
Qy 232 -ERLEGFRGGLGVPLFPPEE---LPGREAAFGVRSRPLGEVR-----HALTHR 276  
Db 248 QRPSSGLMGLGVCPQGADEESLRQWLAQRIADNLTQ-LTAEFRHFSHFHLDIVPWL 306  
Qy 281 -----EVRLGALWEGEGEDP---WKRPPLKLMKYLKRALPL 313  
Db 307 PVSSFTGCMDEGNALWYNLAQPSVGLAAYVERLQO-LRTGAPV 350

RESULT 6  
US-09-651-656-37  
; Sequence 37, Application US/09651656  
; Patent No. 6340566  
; GENERAL INFORMATION:  
; APPLICANT: MCCUTHEN-MALONEY, SANDRA  
; APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY  
; TITLE OF INVENTION: DETECTION AND QUANTIFICATION OF SINGLE NUCLEOTIDE  
; TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,

;; TITLE OF INVENTION: DNA DAMAGE AND DNA MISMATCHES  
;; FILE REFERENCE: IL-10689

;; CURRENT APPLICATION NUMBER: US/09/651,656

;; CURRENT FILING DATE: 2000-08-29

;; PRIOR APPLICATION NUMBER: 60/192,764

;; PRIOR FILING DATE: 2000-03-28

;; NUMBER OF SEQ ID NOS: 106

;; SOFTWARE: Patent In Ver. 2.1

;; SEQ ID NO: 37

;; LENGTH: 350

;; TYPE: PRT

;; ORGANISM: Escherichia coli

;; US-09-651-656-37

Query Match

Best Local Similarity 27.7%; Score 471.5; DB 4; Length 350;

Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAAYRENAR-PLPMRGKNDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66

DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVYLQOTVATVYFERFMARFPTVLDLANAP 69

QY 67 LEEVLYRWOGAGYRRAEHLHRLARSVEEL-----PSPFAELRGLPGGPTTAASIA 121

DB 70 LDEVLYHMTGLGYAARNLHKAQOVATLHGKRPETFEFEVVALPGVGRSTAGATILSLS 129

QY 122 FGERVAADGNVRRVLSRLFARES--PKRE-----LFALAQCLLP-GVDEGVNQALMEL 175

DB 130 LGHNFPLDGNVRYLARCAYASGMPKKEVENKIMLSIEQVPAVER--FNQAMMDL 187

QY 176 GATVCLPKPRCGACPL--GAFRCRKEAPGRYPARRRRAKEERLVALVLLGKRGVHL-- 231

DB 188 GAMICTRSKPKSLCPDLONGCIAANNMSMALYPGKKPKQPLPERTGYFLLQHEDEVLLA 247

QY 232 -ERLEGRFOGLYGVLPFPEE-----LPGREAFGVSRPLGVRALTHRLRY----- 280

DB 248 QRPSPGLMGGLYCFQFADDESLNQWLAQRIADNLTO-LTAFRHTFESHFLDIYPMWL 306

QY 281 -----EVRGALMBEGEDP---WKRPPLKMEKVLKALPL 313

DB 307 PVSSFTGCMDEGNALMTNLNLAQPSVGLAABVERLLOO-LRTGAPV 350

Query Match

Best Local Similarity 27.7%; Score 471.5; DB 4; Length 350;

Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAAYRENAR-PLPMRGKNDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66

DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVYLQOTVATVYFERFMARFPTVLDLANAP 69

QY 67 LEEVLYRWOGAGYRRAEHLHRLARSVEEL-----PSPFAELRGLPGGPTTAASIA 121

DB 70 LDEVLYHMTGLGYAARNLHKAQOVATLHGKRPETFEFEVVALPGVGRSTAGATILSLS 129

QY 122 FGERVAADGNVRRVLSRLFARES--PKRE-----LFALAQCLLP-GVDEGVNQALMEL 175

DB 130 LGHNFPLDGNVRYLARCAYASGMPKKEVENKIMLSIEQVPAVER--FNQAMMDL 187

QY 176 GATVCLPKPRCGACPL--GAFRCRKEAPGRYPARRRRAKEERLVALVLLGKRGVHL-- 231

DB 188 GAMICTRSKPKSLCPDLONGCIAANNMSMALYPGKKPKQPLPERTGYFLLQHEDEVLLA 247

QY 232 -ERLEGRFOGLYGVLPFPEE-----LPGREAFGVSRPLGVRALTHRLRY----- 280

DB 248 QRPSPGLMGGLYCFQFADDESLNQWLAQRIADNLTO-LTAFRHTFESHFLDIYPMWL 306

QY 281 -----EVRGALMBEGEDP---WKRPPLKMEKVLKALPL 313

DB 307 PVSSFTGCMDEGNALMTNLNLAQPSVGLAABVERLLOO-LRTGAPV 350

Query Match

Best Local Similarity 27.7%; Score 471.5; DB 4; Length 350;

Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAAYRENAR-PLPMRGKNDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66

DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVYLQOTVATVYFERFMARFPTVLDLANAP 69

QY 67 LEEVLYRWOGAGYRRAEHLHRLARSVEEL-----PSPFAELRGLPGGPTTAASIA 121

DB 70 LDEVLYHMTGLGYAARNLHKAQOVATLHGKRPETFEFEVVALPGVGRSTAGATILSLS 129

QY 122 FGERVAADGNVRRVLSRLFARES--PKRE-----LFALAQCLLP-GVDEGVNQALMEL 175

DB 130 LGHNFPLDGNVRYLARCAYASGMPKKEVENKIMLSIEQVPAVER--FNQAMMDL 187

QY 176 GATVCLPKPRCGACPL--GAFRCRKEAPGRYPARRRRAKEERLVALVLLGKRGVHL-- 231

DB 188 GAMICTRSKPKSLCPDLONGCIAANNMSMALYPGKKPKQPLPERTGYFLLQHEDEVLLA 247

QY 232 -ERLEGRFOGLYGVLPFPEE-----LPGREAFGVSRPLGVRALTHRLRY----- 280

DB 248 QRPSPGLMGGLYCFQFADDESLNQWLAQRIADNLTO-LTAFRHTFESHFLDIYPMWL 306

QY 281 -----EVRGALMBEGEDP---WKRPPLKMEKVLKALPL 313

DB 307 PVSSFTGCMDEGNALMTNLNLAQPSVGLAABVERLLOO-LRTGAPV 350

Query Match

Best Local Similarity 27.7%; Score 471.5; DB 4; Length 350;

Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAAYRENAR-PLPMRGKNDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66

DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVYLQOTVATVYFERFMARFPTVLDLANAP 69

QY 67 LEEVLYRWOGAGYRRAEHLHRLARSVEEL-----PSPFAELRGLPGGPTTAASIA 121

DB 70 LDEVLYHMTGLGYAARNLHKAQOVATLHGKRPETFEFEVVALPGVGRSTAGATILSLS 129

QY 122 FGERVAADGNVRRVLSRLFARES--PKRE-----LFALAQCLLP-GVDEGVNQALMEL 175

DB 130 LGHNFPLDGNVRYLARCAYASGMPKKEVENKIMLSIEQVPAVER--FNQAMMDL 187

QY 176 GATVCLPKPRCGACPL--GAFRCRKEAPGRYPARRRRAKEERLVALVLLGKRGVHL-- 231

DB 188 GAMICTRSKPKSLCPDLONGCIAANNMSMALYPGKKPKQPLPERTGYFLLQHEDEVLLA 247

QY 232 -ERLEGRFOGLYGVLPFPEE-----LPGREAFGVSRPLGVRALTHRLRY----- 280

DB 248 QRPSPGLMGGLYCFQFADDESLNQWLAQRIADNLTO-LTAFRHTFESHFLDIYPMWL 306

QY 281 -----EVRGALMBEGEDP---WKRPPLKMEKVLKALPL 313

DB 307 PVSSFTGCMDEGNALMTNLNLAQPSVGLAABVERLLOO-LRTGAPV 350

Query Match

Best Local Similarity 27.7%; Score 471.5; DB 4; Length 350;

Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAAYRENAR-PLPMRGKNDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66

DB 10 VLDWYDKYGRKTLPMQIDKTPYKWLSEVYLQOTVATVYFERFMARFPTVLDLANAP 69

QY 67 LEEVLYRWOGAGYRRAEHLHRLARSVEEL-----PSPFAELRGLPGGPTTAASIA 121

DB 70 LDEVLYHMTGLGYAARNLHKAQOVATLHGKRPETFEFEVVALPGVGRSTAGATILSLS 129

QY 122 FGERVAADGNVRRVLSRLFARES--PKRE-----LFALAQCLLP-GVDEGVNQALMEL 175

DB 130 LGHNFPLDGNVRYLARCAYASGMPKKEVENKIMLSIEQVPAVER--FNQAMMDL 187

QY 176 GATVCLPKPRCGACPL--GAFRCRKEAPGRYPARRRRAKEERLVALVLLGKRGVHL-- 231

DB 188 GAMICTRSKPKSLCPDLONGCIAANNMSMALYPGKKPKQPLPERTGYFLLQHEDEVLLA 247

QY 232 -ERLEGRFOGLYGVLPFPEE-----LPGREAFGVSRPLGVRALTHRLRY----- 280

DB 248 QRPSPGLMGGLYCFQFADDESLNQWLAQRIADNLTO-LTAFRHTFESHFLDIYPMWL 306

QY 281 -----EVRGALMBEGEDP---WKRPPLKMEKVLKALPL 313

DB 307 PVSSFTGCMDEGNALMTNLNLAQPSVGLAABVERLLOO-LRTGAPV 350

Query Match

Best Local Similarity 26.1%; Score 443; DB 4; Length 360;

Matches 103; Conservative 56; Mismatches 119; Indels 22; Gaps 7;

QY 3 AMRKALLAWYRENARPLPMRGKNDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKAL 62

DB 19 SFKKDIDBWHKNOBMPMRETTNPPYILSEVMLOQOVNTIDYVYRFTHRPTIOSL 78

QY 63 AASLEEVLYRWOGAGYRRAEHLHRLARSVE-----ELPSPFAELRGLPGGYTAAV 117

DB 79 SEANEDDEVLYHMTGLGYAARNLHKAQOVATLHGKRPETFEFEVVALPGVGRSTAGATILSLS 129

QY 118 ASTAFGERVAADGNVRRVLSRLFARES--PKRE-----LFALAQCLLP-GVDEGVNQALMEL 174

DB 139 MSAFHNHPLATYDGNVRYLARCAYASGMPKKEVENKIMLSIEQVPAVER--FNQAMMDL 187

QY 175 LGATVCLPKPRCGACPL--GAFRCRKEAPGRYPARRRRAKEERLVALVLLGKRGVHL-- 231

DB 199 LGALVCTPKSPCLCPDLONGCIAANNMSMALYPGKKPKQPLPERTGYFLLQHEDEVLLA 247

QY 232 -ERLEGRFOGLYGVLPFPEE-----LPGREAFGVSRPLGVRALTHRLRY----- 280

DB 259 LKRRREKLLNGKMOFPM--REQTANADVISDDGKSTIETINERVFVKLQHFTHLTWEIKV 316

Query Match

Best Local Similarity 26.1%; Score 443; DB 4; Length 360;

Matches 103; Conservative 56; Mismatches 119; Indels 22; Gaps 7;

QY 3 AMRKALLAWYRENARPLPMRGKNDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKAL 62

DB 19 SFKKDIDBWHKNOBMPMRETTNPPYILSEVMLOQOVNTIDYVYRFTHRPTIOSL 78

QY 63 AASLEEVLYRWOGAGYRRAEHLHRLARSVE-----ELPSPFAELRGLPGGYTAAV 117

DB 79 SEANEDDEVLYHMTGLGYAARNLHKAQOVATLHGKRPETFEFEVVALPGVGRSTAGATILSLS 129

QY 118 ASTAFGERVAADGNVRRVLSRLFARES--PKRE-----LFALAQCLLP-GVDEGVNQALMEL 174

DB 139 MSAFHNHPLATYDGNVRYLARCAYASGMPKKEVENKIMLSIEQVPAVER--FNQAMMDL 187

QY 175 LGATVCLPKPRCGACPL--GAFRCRKEAPGRYPARRRRAKEERLVALVLLGKRGVHL-- 231

DB 199 LGALVCTPKSPCLCPDLONGCIAANNMSMALYPGKKPKQPLPERTGYFLLQHEDEVLLA 247

QY 232 -ERLEGRFOGLYGVLPFPEE-----LPGREAFGVSRPLGVRALTHRLRY----- 280

DB 259 LKRRREKLLNGKMOFPM--REQTANADVISDDGKSTIETINERVFVKLQHFTHLTWEIKV 316

Query Match

Best Local Similarity 26.1%; Score 443; DB 4; Length 360;

Matches 103; Conservative 56; Mismatches 119; Indels 22; Gaps 7;

QY 3 AMRKALLAWYRENARPLPMRGKNDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKAL 62

DB 19 SFKKDIDBWHKNOBMPMRETTNPPYILSEVMLOQOVNTIDYVYRFTHRPTIOSL 78

QY 63 AASLEEVLYRWOGAGYRRAEHLHRLARSVE-----ELPSPFAELRGLPGGYTAAV 117

DB 79 SEANEDDEVLYHMTGLGYAARNLHKAQOVATLHGKRPETFEFEVVALPGVGRSTAGATILSLS 129

QY 118 ASTAFGERVAADGNVRRVLSRLFARES--PKRE-----LFALAQCLLP-GVDEGVNQALMEL 174

DB 139 MSAFHNHPLATYDGNVRYLARCAYASGMPKKEVENKIMLSIEQVPAVER--FNQAMMDL 187

QY 175 LGATVCLPKPRCGACPL--GAFRCRKEAPGRYPARRRRAKEERLVALVLLGKRGVHL-- 231

DB 199 LGALVCTPKSPCLCPDLONGCIAANNMSMALYPGKKPKQPLPERTGYFLLQHEDEVLLA 247

QY 232 -ERLEGRFOGLYGVLPFPEE-----LPGREAFGVSRPLGVRALTHRLRY----- 280

DB 259 LKRRREKLLNGKMOFPM--REQTANADVISDDGKSTIETINERVFVKLQHFTHLTWEIKV 316

Query Match

Best Local Similarity 26.1%; Score 443; DB 4; Length 360;

Matches 103; Conservative 56; Mismatches 119; Indels 22; Gaps 7;

QY 3 AMRKALLAWYRENARPLPMRGKNDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKAL 62

DB 19 SFKKDIDBWHKNOBMPMRETTNPPYILSEVMLOQOVNTIDYVYRFTHRPTIOSL 78

QY 63 AASLEEVLYRWOGAGYRRAEHLHRLARSVE-----ELPSPFAELRGLPGGYTAAV 117

DB 79 SEANEDDEVLYHMTGLGYAARNLHKAQOVATLHGKRPETFEFEVVALPGVGRSTAGATILSLS 129

QY 118 ASTAFGERVAADGNVRRVLSRLFARES--PKRE-----LFALAQCLLP-GVDEGVNQALMEL 174

DB 139 MSAFHNHPLATYDGNVRYLARCAYASGMPKKEVENKIMLSIEQVPAVER--FNQAMMDL 187

QY 175 LGATVCLPKPRCGACPL--GAFRCRKEAPGRYPARRRRAKEERLVALVLLGKRGVHL-- 231

DB 199 LGALVCTPKSPCLCPDLONGCIAANNMSMALYPGKKPKQPLPERTGYFLLQHEDEVLLA 247

QY 232 -ERLEGRFOGLYGVLPFPEE-----LPGREAFGVSRPLGVRALTHRLRY----- 280

DB 259 LKRRREKLLNGKMOFPM--REQTANADVISDDGKSTIETINERVFVKLQHFTHLTWEIKV 316

Query Match

Best Local Similarity 26.1%; Score 443; DB 4; Length 360;

Matches 103; Conservative 56; Mismatches 119; Indels 22; Gaps 7;

QY 3 AMRKALLAWYRENARPLPMRGKNDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKAL 62

DB 19 SFKKDIDBWHKNOBMPMRETTNPPYILSEVMLOQOVNTIDYVYRFTHRPTIOSL 78

QY 63 AASLEEVLYRWOGAGYRRAEHLHRLARSVE-----ELPSPFAELRGLPGGYTAAV 117

DB 79 SEANEDDEVLYHMTGLGYAARNLHKAQOVATLHGKRPETFEFEVVALPGVGRSTAGATILSLS 129

QY 118 ASTAFGERVAADGNVRRVLSRLFARES--PKRE-----LFALAQCLLP-GVDEGVNQALMEL 174

DB 139 MSAFHNHPLATYDGNVRYLARCAYASGMPKKEVENKIMLSIEQVPAVER--FNQAMMDL 187

QY 175 LGATVCLPKPRCGACPL--GAFRCRKEAPGRYPARRRRAKEERLVALVLLGKRGVHL-- 231

DB 199 LGALVCTPKSPCLCPDLONGCIAANNMSMALYPGKKPKQPLPERTGYFLLQHEDEVLLA 247

QY 232 -ERLEGRFOGLYGVLPFPEE-----LPGREAFGVSRPLGVRALTHRLRY----- 280

DB 259 LKRRREKLLNGKMOFPM--REQTANADVISDDGKSTIETINERVFVKLQHFTHLTWEIKV 316

Query Match

Best Local Similarity 26.1%; Score 443; DB 4; Length 360;

Matches 103; Conservative 56; Mismatches 119; Indels 22; Gaps 7;

QY 3 AMRKALLAWYRENARPLPMRGKNDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKAL 62

GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
TITLE OF INVENTION: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328.352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 4982  
LENGTH: 346  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-4982

Query Match  
Best Local Similarity 33.38; Pred. No. 2,6e-36;  
Matches 113; Conservative 60; Mismatches 118; Indels 48; Gaps 11;

QY 7 ALLAWYRENNAR-PPMGEKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAA 65  
| | | : : : | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 11 ALLWYFOHGRHNDLPQVADDPYKVMYSEIMLQOTQVTVLQYEDREREPTVEALGYA 70  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 66 SLEEVLRVWOGAGYYRRAEHLRLARSVE--LPPSFAELRGLEPGLPYTAATAVASTAF 122  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 71 TWDEVAPYMACLGYYARARNLHKAAGLVAOQCKPPEPLEMIALPGIGRSTAGALMSGL 130  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 123 GERAAADGNVRRVLSLFARES-----PKKEKELFALAQGLPEGVDPGVNQAALMELGA 177  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 131 ROYVINDGKVRKRLARFALIEDLSKPOHEREMWKLAEELCPTRNMD-YTQALMDIGA 189  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 178 TVCLPKRRCGACPLGAFAC-----RGKEAPGRYPAPRRKRAKEERLVALVLGKGVHLER 233  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 190 TICFPKPKCLICXCMQAHCAAYOQGLDELPFKPKPKPKTKTADVLIIQCEDEWFQOR 249  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 234 -LEGRFQGLGVPLPEPPE-ELPGREAAFGVRSRP-LGEVRHATHN----- 275  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 250 QAHGLMGGLFCLPILNENHERLKLQOQKFLQPOQOTFQISHSFHTFWLNAHVHVPD 309  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 276 --RLRVEVRGALMBEGBEDPWKRRLPKLMKVKLKPALP 312  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 310 QKEHLAIELEG-----QMLSP-----EQALAKGPV 334  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |

RESULT 10  
US-08-663-023-17  
Sequence 17, Application US/08663023  
Patent No. 5763178  
GENERAL INFORMATION:  
APPLICANT: CHIRIKIJAN, Jack G.  
APPLICANT: COLLIER, G. Bruce  
TITLE OF INVENTION: OSCILLATING SIGNAL AMPLIFIER FOR NUCLEIC  
TITLE OF INVENTION: ACID DETECTION  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/663.023  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/012.950  
FILING DATE: 06-MAR-1996  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/483.089  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 66669/110  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-663-023-17

Query Match  
Best Local Similarity 31.7%; Pred. No. 1e-19;  
Matches 63; Conservative 40; Mismatches 85; Indels 11; Gaps 4;

QY 8 LLAWYRENNARPLPMGEKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAA 67  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 16 ILTFWNTDRDPPMHRHTDPPVILITTEILRRTTAGHVKKIYDKPFVYKCFEDILKTPK 75  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 68 EEVLRVWOGAGY-YRRAEHLRLARSV-----EELPPSFAELRGLEPGLPYTAATAVASTA 121  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 76 SEIARDIEIGISNORAEQLKELARVINDYGRVRRRKAILLDPGVKTYTCAAVMCLA 135  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 122 FGERAAADGNVRRVLSLFARE-----SPKEKELFALAQGLPEGVDPGVNQAALMELGA 177  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 136 FGKKAAMVDANFVRINNTFEGGSYENLNINIKALMELAEITLVPGKCHD-FNLGLMDISA 194  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
QY 178 TVCLPKRRCGACPLGAFAC 196  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 195 ITCAPRKPCKEKGMSKLC 213  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |

RESULT 11  
US-09-402-959A-10  
Sequence 10, Application US/09402959A  
Patent No. 6548247  
GENERAL INFORMATION:  
APPLICANT: CHIRIKIJAN, Jack G.  
APPLICANT: BAZAR, LEONARD S.  
TITLE OF INVENTION: DETECTION AND MAPPING OF POINT MUTATIONS USING PARTIAL  
FILE REFERENCE: 066669/0118  
CURRENT APPLICATION NUMBER: US/09/402.959A  
CURRENT FILING DATE: 1999-12-09  
PRIOR APPLICATION NUMBER: PCT/US98/06878  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/043,184  
PRIOR FILING DATE: 1997-04-16  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 10  
LENGTH: 221  
TYPE: PRT  
ORGANISM: Unknown Organism  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: ORF 10  
US-09-402-959A-10

Query Match  
Best Local Similarity 31.7%; Pred. No. 1e-19;  
Matches 63; Conservative 40; Mismatches 85; Indels 11; Gaps 4;

QY 8 LLAWYRENNARPLPMGEKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAA 67  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |  
DB 16 ILTFWNTDRDPPMHRHTDPPVILITTEILRRTTAGHVKKIYDKPFVYKCFEDILKTPK 75  
: : : | | | | | | : : : | | | : : : | | | : : : | | | : : : |

QY 68 EVLRVWOGAGY-YRAAEHLRLARSV-----EELPPSPAEIRGLPGYTAAGVASTA 121  
 Db 76 SEIMKDIEIGLSNORAPQELKELARVYINDYGGVFPNRRKALIDLEPVGGYTCGAAMVCLA 135  
 QY 122 FGERAAVADGNVRVRLRFARE-----SPKEKELFALAGLPEGVDPGVNQAELMELCA 177  
 Db 136 FGKAAAVDANFVYVIRKRYFGGSYENLNHNKALMELAEITLVGKCRD-FNLGLMDFSA 194  
 QY 178 TVCLPKRRRCGACPLGAFNC 196  
 Db 195 IICAPRRKCEKCGMSKLC 213

## RESULT 12

US-09-107-532A-3843  
 ; Sequence 3843, Application US/09107532A  
 ; Patent No. 6583275  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 7310  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 ; STREET: 100 Beaver Street  
 ; CITY: Waltham  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02354  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: CD-ROM ISO9660  
 ; COMPUTER: PC  
 ; OPERATING SYSTEM: <Unknown>  
 ; SOFTWARE: ASCII  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/107,532A  
 ; FILING DATE: 30-Jun-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/085,598  
 ; FILING DATE: 14 May 1998  
 ; APPLICATION NUMBER: 60/051571  
 ; FILING DATE: July 2, 1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Arinello, Pamela Deneke  
 ; REGISTRATION NUMBER: 40,489  
 ; REFERENCE/DOCKET NUMBER: GTC-012  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (781)893-5007  
 ; TELEFAX: (781)893-8277  
 ; INFORMATION FOR SEQ ID NO: 3843:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 227 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHEICAL: YES  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Enterococcus faecium  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (8) LOCATION 1...227  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 3843:  
 US-09-107-532A-3843

Query Match 11.1%; Score 188; DB 4; Length 227;  
 Best Local Similarity 30.6%; Pred. No. 5.3e-11;  
 Matches 63; Conservative 36; Mismatches 79; Indels 28; Gaps 9;

QY 26 DEYRVLYSEVLLQ---TRVEQALPYRRFLERPTLKALAASLEEVLRVWOGAGYR- 81  
 Db 31 NPELLLAIVLLSAQATDVSVNKATP---DLFASPTPDALAEASIDELIILKITITGLYNN 87  
 QY 82 RAEHLRLARSVE-----LPPSPAEIRGLPGYTAAGVASTAIGERAAVADGNVRV 136

Db 88 KAKINAKAQQILIERFDGQVPTSRBELMSLPVGRKTAHVVLGDAFGIPALIAVDTHVERV 147  
 QY 137 LSRLEFARSPKEKELFALAGLPEGVDPGVW---NQALMELGATVCLPKRRRCACPLG 193  
 Db 148 SKRL--RICKLDAIVMEVEETLMRK-VPELWVKTHHTLIFGRYHCTARNKCEVPL 204  
 QY 194 AFGR-GKEAPGRYPAPRRKREERL 218  
 Db 205 SICODK-----NRRLKEKAL 221

## RESULT 13

US-08-808-550-36  
 ; Sequence 36, Application US/08808550  
 ; Patent No. 5871992  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Teebo, George W.  
 ; APPLICANT: Hilbert, Timothy P.  
 ; TITLE OF INVENTION: MAMMALIAN ENDONUCLEASE III AND  
 ; NUMBER OF SEQUENCES: 42  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: David A. Jackson, Esq.  
 ; STREET: 411 Hackensack Ave, Continental Plaza, 4th  
 ; CITY: Hackensack  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 07601  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentln Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/808,550  
 ; FILING DATE: 26-FEB-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jackson Esq., David A.  
 ; REGISTRATION NUMBER: 26,742  
 ; REFERENCE/DOCKET NUMBER: 1049-1-001 N  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 201-487-5800  
 ; TELEFAX: 201-343-1684  
 ; INFORMATION FOR SEQ ID NO: 36:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 207 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHEICAL: NO  
 ; FRAGMENT TYPE: N-terminal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Escherichia coli  
 US-08-808-550-36

Query Match 10.4%; Score 176; DB 2; Length 207;  
 Best Local Similarity 32.0%; Pred. No. 7.6e-10;  
 Matches 58; Conservative 33; Mismatches 70; Indels 20; Gaps 8;

QY 28 YRVLYSEVLLQOTRVEQALPYRRFLERPTLKALAASLEEVLRVWOGAGYR-RAEHL 86  
 Db 29 FOVLVAIWLSSQTRDEVAAAMKRLKDHGSLGKILFEKVPDETLLCVCVGFKKRAVYL 88  
 QY 87 HRLARSVE-----LPPSPAEIRGLPGYTAAGVASTAIGERAAVADGNVRVLSRL 140  
 Db 89 OKTKAILKDDSGGIPIPSLDGLCALPEVGGPKMANLWQJAGECVGIADVTHHRTSNRL 148  
 QY 141 -FARSPKEKELFALAGLPEGVDPGVW---NQALMELGATVCLPKRRRCACPLGAFNC 196



Db 149 GWKSTPEKTKAL-EILLPK-----SEWQPINHLIVFGWQCPVPRKCGTC---LC 199

Qy 197 R 197

Db 200 R 200

```

1      RESULT 14
2      US-08-808-550-39
3      Sequence 39, Application US/08800550
4      Patent No. 5871992
5      GENERAL INFORMATION:
6      APPLICANT: Teebor, George W.
7      APPLICANT: Hilbert, Timothy P.
8      TITLE OF INVENTION: MAMMALIAN ENDONUCLEASE III AND
9      TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF
10     NUMBER OF SEQUENCES: 42
11     CORRESPONDENCE ADDRESS:
12     ADDRESSEE: David A. Jackson, Esq.
13     STREET: 411 Hackensack Ave, Continental Plaza, 4th
14     STREET: floor
15     CITY: Hackensack
16     STATE: New Jersey
17     COUNTRY: USA
18     ZIP: 07601
19     COMPUTER READABLE FORM:
20     MEDIUM TYPE: Floppy disk
21     COMPUTER: IBM PC compatible
22     OPERATING SYSTEM: PC-DOS/MS-DOS
23     SOFTWARE: Patentin Release #1.0, Version #1.30
24     CURRENT APPLICATION DATA:
25     APPLICATION NUMBER: US/08/808,550
26     FILING DATE: 26-FEB-1997
27     CLASSIFICATION: 435
28     ATTORNEY/AGENT INFORMATION:
29     NAME: Jackson Esq., David A.
30     REGISTRATION NUMBER: 26,742
31     REFERENCE/DOCKET NUMBER: 1049-1-001 N
32     TELECOMMUNICATION INFORMATION:
33     TELEPHONE: 201-487-5800
34     TELEFAX: 201-343-1684
35     INFORMATION FOR SEQ ID NO: 39:
36     SEQUENCE CHARACTERISTICS:
37     LENGTH: 259 amino acids
38     TYPE: amino acid
39     STRANDEDNESS: single
40     TOPOLOGY: linear
41     MOLECULE TYPE: protein
42     HYPOTHETICAL: NO
43     FRAGMENT TYPE: N-terminal
44     ORIGINAL SOURCE:
45     ORGANISM: C. elegans
46     US-08-808-550-39

```

Query Match	10.4%	Score 176;	DB 2;	Length 259;
Best Local Similarity	32.0%	Pred. No. 1e-09;		
Matches	58;	Conservative	33;	Mismatches 70;
				Indels 20;
				Gaps 8

```

1      RESULT 15
2      US-08-808-550-31
3      ; Sequence 31, Application US/08080550
4      ; Patent No. 5671992
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Teebort, George W.
7      ; APPLICANT: Hilbert, Timothy P.
8      ; TITLE OF INVENTION: MAMMALIAN ENDONUCLEASE III AND
9      ; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF
10     ; NUMBER OF SEQUENCES: 42
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: David A. Jackson, Esq.
13     ; STREET: 411 Hackensack Ave, Continental Plaza, 4th
14     ; STREET: Floor
15     ; CITY: Hackensack
16     ; STATE: New Jersey
17     ; COUNTRY: USA
18     ; ZIP: 07601
19     ; COMPUTER READABLE FORM:
20     ; MEDIUM TYPE: Floppy disk
21     ; COMPUTER: IBM PC compatible
22     ; OPERATING SYSTEM: PC-DOS/MS-DOS
23     ; SOFTWARE: Patentin Release #1.0, Version #1.30
24     ; CURRENT APPLICATION DATA:
25     ; APPLICATION NUMBER: US/08/808,550
26     ; FILING DATE: 26-FEB-1997
27     ; CLASSIFICATION: 435
28     ; ATTORNEY/AGENT INFORMATION:
29     ; NAME: Jackson Esq., David A.
30     ; REGISTRATION NUMBER: 26,742
31     ; REFERENCE/DOCKET NUMBER: 1049-1-001 N
32     ; TELECOMMUNICATION INFORMATION:
33     ; TELEPHONE: 201-487-5800
34     ; TELEFAX: 201-343-1684
35     ; INFORMATION FOR SEQ ID NO: 31:
36     ; SEQUENCE CHARACTERISTICS:
37     ; LENGTH: 209 amino acids
38     ; TYPE: amino acid
39     ; STRANDEDNESS: single
40     ; TOPOLOGY: linear
41     ; MOLECULE TYPE: protein
42     ; HYPOTHETICAL: NO
43     ; FRAGMENT TYPE: N-terminal
44     ; ORIGINAL SOURCE:
45     ; ORGANISM: Bacillus subtilis
46     ; US-08-808-550-31

```

Query Match	10.1%	Score 172	DB 2	Length 209
Best Local Similarity	28.1%	Pred No. 1.9e-09		
Matches	56	Conservative	30	Mismatches 65
				Indels 48
				Gaps 6

Search completed: September 15, 2003, 13:54:22  
Job time : 32 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 15, 2003, 13:48:09 ; Search time 40 Seconds  
(without alignments)  
781.371 Million cell updates/sec

Title: US-09-938-901-2  
Perfect score: 1700  
Sequence: 1 VEAMRKALLAWYRENARPLP.....VLKRALPLLAHGVPLPDA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	669.5	39.4	363	2	A75294 A/G-specific aden
2	551.5	32.4	375	2	AD3432 A/G-specific aden
3	520	30.6	349	2	H87295 A/G-specific aden
4	515.5	30.3	367	2	AH2674 A/G-specific aden
5	515.5	30.3	367	2	F97456 A/G-specific aden
6	503	29.6	349	2	B81855 adenine glycosylas
7	501.5	29.5	369	2	A69802 A/G-specific aden
8	498	29.3	349	2	D81088 A/G-specific aden
9	497.5	29.3	372	2	C83766 adenine glycosylas
10	492	28.9	387	2	A86730 purine nucleosidas
11	491.5	28.9	350	2	AG0879 A/G-specific aden
12	491	28.9	371	2	AH0116 A/G-specific aden
13	484.5	28.5	350	2	A40647 methylation-indepe
14	483	28.4	365	2	AD1657 A/G-specific aden
15	478.5	28.1	353	2	D82320 A/G-specific aden
16	478	28.1	365	2	A11285 A/G-specific aden
17	477.5	28.1	355	2	G83003 A / G specific ade
18	477	28.1	369	2	H86540 adenine glycosylas
19	477	28.1	369	2	A72082 adenine glycosylas
20	476	28.0	339	2	B81258 A/G-specific aden
21	471.5	27.7	350	2	B38535 A/G-specific aden
22	469.5	27.6	350	2	E91108 adenine glycosylas
23	469.5	27.6	350	2	H85953 adenine glycosylas
24	467	27.5	345	2	C89974 hypothetical prote
25	459.5	27.0	381	2	E95142 A/G-specific aden
26	459.5	27.0	391	2	C98010 A/G-specific aden
27	458.5	27.0	369	2	B71557 probable adenine g
28	445	26.2	378	2	C64091 adenine glycosylas
29	442.5	26.0	461	2	T43679 A/G-specific aden

30	442.5	26.0	461	2	T38390 A/G-specific aden
31	441.5	26.0	371	2	A81709 A/G-specific aden
32	436	25.6	357	2	D82624 A/G-specific aden
33	423	24.9	328	2	F64537 A/G-specific aden
34	420	24.7	328	2	E71969 A/G-specific aden
35	414	24.4	312	2	G84305 hypothetical prote
36	379	22.3	308	2	T36366 probable adenine g
37	378.5	22.3	350	2	D84994 A/G-specific aden
38	374.5	22.0	608	2	T06632 hypothetical prote
39	365.5	21.5	304	2	F70804 probable DNA glyco
40	360	21.2	297	2	C87149 probable DNA glyco
41	342	20.1	277	2	F71335 probable A/G-speci
42	313.5	18.4	286	2	A72682 probable A/G-speci
43	305.5	18.0	215	2	H97084 A/G-specific DNA g
44	274.5	16.1	221	2	S30312 endonuclease III b
45	227	13.4	213	2	F72387 endonuclease III -

ALIGNMENTS

RESULT 1

A75294  
A/G-specific adenine glycosylase - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 08-Dec-2000  
C:Accession: A75294  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: A75294  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-363 <WHI>  
A:Cross-references: GB:AE002060; GB:AE00513; NID:96460082; PIDN:AP11831.1; PID:9646  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR2285  
A:Map position: 1  
C:Superfamily: A/G-specific adenine glycosylase

Query Match	39.4%	Score 669.5;	DB 2;	Length 363;
Best Local Similarity	46.2%	Pred. No. 9.7e-45;		
Matches 154;	Conservative 48;	Mismatches 96;	Indels 35;	Gaps 11;
QY	1	VEAMRKALLAWYRENARPLPWR----	GEKDPYRYLSEVLLQOTRVFOALPYRRLERF	56
DB	17	VGALRDLGMPDRAGDLPWRUGDBGRDPRVWVAETLLQOTQVARGSLYERLEAF	76	
QY	57	PTLKALAAASLEUEYLWQAGAGYRRAEHLRLARSVEE--LPPSFALRGLPGLGYTA	114	
DB	77	PTYQALAAAPQDAVLKAMECGCYTARARNLHRAAIIIDEGGFQPDYAGWMLALPEGVYTA	136	
QY	115	AAVASIAFGERVAADGNVRVLSRLFARESPEKELFALAQGLLPBGVDGWNQALME	174	
DB	137	AAVSSIALGEPRAVNDGNVRVLSRLFAEHPSPDKWQEQDADRLD-DPARPGANNEAVMD	195	
QY	175	LGATVCLPKRPRGACGLGAFCKRKE--AGRGYAPARR-RAKERLVALVLGRKVVHL	231	
DB	196	LGATICVPSKPACDRCPVSAHCAAYQGGPDPAPAPARQAREVRVALDIGAEYAVL	255	
QY	232	ERLEGR-FQGLGYPLRPPEELPGRENAFVRSR-----PLGEVHAHTHRLRL	279	
DB	256	EKREGSLGLFLGFLPL--BEIGARETAADALARLQARLGAEVKECCIGTQHGHTHRLS	312	
QY	280	VEVRGALMEGEDDPWKRP-----LPRIMEKVL	307	
DB	313	VEV-----YRAEDRP-RQPVGCAALSLRLDKAL	340	

[illegible]

```

Query Match          30.68; Score 520; DB 2; Length 349;
Best Local Similarity 40.68; Pred. No. 4 2e-33;
Matches 141; Conservative 33; Mismatches 137; Indels 36; Gaps 10;

QY 2 EAMRKALLAWYRENAARPLPW-----GKDPYRVLYSEVLLQOTRVEQALPYRRF 52
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5 DALRSALLAWYDQAORDLARVGAPEARRAGVSRSDPYRWLSEVWLQOTVPNATPFELSF 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 TORMPYVLVDLAAYVEDDGLMAMAGLGYAARNLILACARVANDHGDGVFPTEBGLALP 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 53 LEERPTLKALAAASLEEVLRVNOGAGYRRARLHLRLARVEE-----LRPSAEIRGLP 107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 TORMPYVLVDLAAYVEDDGLMAMAGLGYAARNLILACARVANDHGDGVFPTEBGLALP 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 108 GLEPPTAAAAAASIAFGERVAADVGNVRRVLSRLFARESF---KEKELFALAOGLLPBGVD 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 125 GVGATYTAATAAIALIFDRAANAVVDGNVERVMSRLFAVEAPMPDSKPELKALAGDLVTD-D-R 183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 165 PGVWNOALMELGATVCLPKPRRCGACPLGAFCSG--KEAGRRYPAPRRKRAKEERL-VAL 221
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 PGDMAQALMDLIGATVCKPKRPLCDRCQVSLMCAAYVGAPEETYPKTKKADRRPRHGVAY 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 222 VLLGRKGVHLEERL--EGRFQGLYGVPL-----FPPEELPGRBAAFGVSRPLGVEERHA 272
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 244 VLRGDEVALYVRPPKCLLGGMLGLPTSDMKRTASDYDAEVA-AAPLAAMVROLGAEVH 302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 273 LTHRRRLREVRGALMEGEGEDPWK----RPLKLMKVLKALPL 314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 FTHFSILTLRVFADGANDGDVFWTPREGTLALPSVFLKAMAAORLL 349
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
AH2674
A/G-specific adenine glycosylase muty [Imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence__revision 11-Jan-2002 #text__change 18-Nov-2002
C:Accession: AH2674
R:Wood, D.W.; Seubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo-
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutysavin, T.; Levy, R.; Li, M.; McCl-
e, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam-
ster, E.W.
A:Title: The genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Accession number: AB2577; MUID:21608550; PMID:11743193
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-367 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL41814.1; PID:g17739170; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: muty
A:Map position: circular chromosome
C:Superfamily: A/G-specific adenine glycosylase

Query Match          30.38; Score 515.5; DB 2; Length 367;
Best Local Similarity 37.18; Pred. No. 1e-32;
Matches 130; Conservative 55; Mismatches 124; Indels 41; Gaps 13;

QY 3 AMRKALLAWYRENAARPLPW-----RGER-DPYRVLYSEVLLQOTRVEQALPYRRFL 53
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9 SYAOMLAWYDRHHRRELPMRTSPAMARGRADRYHWLSEVWLQOTTVAVNRPYFLKFL 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 54 ERPPPTLKALAAASLEEVLRVNOGAGYRRARLHLRLARVEE-----LRPSAEIRGLP 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69 AAMPFRVDLTLFAVEDVMAAMAGLGTIYARARNLKCCAEAVAREHGVGFPTDEGLKPLG 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 109 LGEPTAAVAASIAFGERVAADVGNVRRVLSRLFARESF---KEKELFALAOGLLPBGVD 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 IGDYTSAAVAALIFNROAAVMDGNVERVISRLFAIDAPLPGSGPAMAKAKVAELTFP-AERP 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 GVNNOALMELGATVCLPKPRRCGACPLGAFCSG--KEAGRRYPAPRRKRAKEERL-VALY 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 GDAQAAMMDIGATVCTPKRPACALCPNGACALTRDEPRFPVKAANKKPVNPLGAFAV 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

A:Molecule type: DNA
A:Residues: 1-349 >(PAR>
A:Cross-references: GB:A162756; GB:A157959; NID:g7380091; PIDN:CA884842.1; PID:g738
A:Experimental source: serogroup A, strain 82491
C:Genetics:
A:Gene: muty; NMAL614
C:Superfamily: A/G-specific adenine glycosylase
C:Keywords: glycosidase; hydrolase

Query Match          29.6%; Score 503; DB 2; Length 349;
Best Local Similarity 39.7%; Pred. No. 8,9e-32;
Matches 131; Conservative 52; Mismatches 111; Indels 36; Gaps 14;

QY      8 LLANYRENAR-PLEMRGEKBDPYRVLSEVLLQOTRVEDQLPYRRFLEREPTEKLALAAS 66
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB       14 LIRWOKGGRHNHLRPMO-VKNPYSWLSEIMLOQTAVTLDYDYPREFLEKEPTVOTGLAAP 72
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

QY      67 LEVLRLVWQSGAGGYRRRAENHLRLARSVE-----LRPSFAELRLPLGLGPTAAVASIA 121
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB       73 QDEVLSTWAGISGYYSRKARNLHKAAQQVVEQFGTFPEBRKDLETLCGGGRSTAACAIFS 132
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

QY      122 FGEVVAAVDGVVRVLSRLFARE-SPKRKE---LPLAQGLP-EGVDPCVNMQALMEL 175
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB       133 FNRETTIIDGVKKRCPLCVFARDGNPDQKKFKENSMTLAESLLRPSENADMPATVQGIMDL 192
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

QY      176 GATVCCLPRPRPGCGCPGLGAFCRGKEADRGYRAPRRRAAKE---RLVALVYLGRKG-VHL 231
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB       193 GATVCKKRKRPCLCHCCPMADICEAKKQNRTHALEPRKKTAAAEVPTLPYLVWLYVRNDGAILL 252
        | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

QY      232 ER--LEGHFOGLYVPLFPPEELPGRE---AAGSVRSRPLGE---VRHALTHRLRLEVVR 283

```

Db 311 EAQMPSESPDGIWIKRAHLKDYGLPPLPLE 340

RESULT 7  
A69802  
A/G-specific adenine glycosylase homolog yfhg - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 08-Dec-2000  
C:Accession: A69802  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ben-  
C: Bron, S.; Bröhlert, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A: Ehrlich, S.D.; Emerson, P.T.; Entlan, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal-  
lech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.;  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinon-  
A:Authors: Lauder, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau,  
Y.M.; Ogawa, K.; Ogward, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portele-  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanli-  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se-  
akuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Toignoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: A69802  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1569 <KUN>  
A:Cross-references: GB:A29108; GB:AL009126; NID:g26303055; PIDN:CAB12691.1; PID:e11828  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yfhg  
C:Superfamily: A/G-specific adenine glycosylase

Query Match 29.5% Score 501.5; DB 2; Length 369;  
Best Local Similarity 36.1% Pred. No. 1.2e-31;  
Matches 112; Conservative 47; Mismatches 108; Indels 43; Gaps 6;

QY 1 VEAMRKALLAMWYRENNAPLPWGEKDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLK 60  
 Db 13 IQOFRDLLISMFEREQGVLPWRDODPYKVMWSEVMLQOTRVEVLPYRFLERFPTLYE 72  
 QY 61 ALAAASLEEVLRVWOGAGYRRAEHLRLARSVE-----LPPSEAEELRGPLGCPYTA 115  
 Db 73 ALADADEEKVLMKEGIGYYSRRNLQSAVKEVGEYGVLPDEKDFGGLKGVCPYTKG 132  
 QY 116 AVASIAFGERVAADVGNVRVLSRLFA---RESPKEKEFLAAGLLGEGVPGVWNOA 171  
 Db 133 AVLSIAVNNKRIPAVDGNVMWVMSRIISINDIAKPTTRITFEDAIATFISKEKSEFNOG 192  
 QY 172 LMELGATVCLPKPRRCGACPLGAFCRGKEAPGRYPAPRRKRAKE---ERLVALVLGRKG 228  
 Db 193 LMELGALICPKSPSCILCPVQOHCSAFEGTERELPVKSKKKPGIKTMAALVILDEDG 252  
 QY 229 ---VHLERL--EGRFGGLVGPPLFPPEELPGREAFGR-----SRP 265  
 Db 253 QVYIHKRPKSKGLLANLM-----EPPMLETQKGIKTEREDLIAELENEYGIADISDL 304  
 QY 266 LGEVRHALTH 275  
 Db 305 QGVVEHVETH 314

## RESULT 8

A/G-specific adenine glycosylase NMB1396 [Imported] - *Neisseria meningitidis* (strain MCS D81088)  
 C:Species: *Neisseria meningitidis*  
 C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
 R:Retcelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Hart, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; ri, H.; Qin, H.; Yamathavan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Pizza, M. Science 287, 1809-1815, 2000  
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappunli, R.; Ve A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MCS8.  
 A:Reference number: AB1000; MUID:20175755; PMID:10710307  
 A:Accession: D81088  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-349 <STO>  
 A:Cross-references: GB:AE002488; GB:AE002098; NID:g7226631; PIDN:AAFA1760.1; PID:g722663  
 A:Experimental source: serogroup B, strain MCS8  
 C:Genetics:  
 A:Gene: NMB1396  
 C:Superfamily: A/G-specific adenine glycosylase

Query Match 29.3%; Score 498; DB 2; Length 349;  
 Best Local Similarity 39.4%; Pred. No. 2,2e-31;  
 Matches 130; Conservative 52; Mismatches 112; Indels 36; Gaps 14;

QY 8 LLAWYRENNAPLPWGEKDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66  
 Db 14 IIRMQKHGRHRLPMQ-VKNPYCWLSEIMLQOTRVAIVLQYPRFLEKFPVQTLAAP 72  
 QY 67 LEEVLRVWOGAGYRRAEHLRLARSVE-----LPPSEAEELRGPLGCPYTA 121  
 Db 73 QDEVLSLWAGLGYSRRNLHKAQOVVROFGTFPSEKRLDETLCGVGRSTAAICAFS 132  
 QY 122 FGERVAADVGNVRVLSRLFARE-SPEKE---LEFALAGLLP-EGVDPGVWNOALMEL 175  
 Db 133 FNRRTIILIDGNVKKVLCVFAKRDGNPOKKRENSLMTAESLLPSENDMDPAYTOGLMDL 192  
 QY 176 GATVCLPKPRRCGACPLGAFCRGKEAPGRYPAPRRKRAKE---RLVALVLGRKG-VHL 231  
 Db 193 GATVCKRTKPLCHOCPMADICEAKQNTALPKRTAAEAEPVLPVLYLRNDDGATLL 252  
 QY 232 ER--LEGRFGGLVGPPLFPPEELGRF---AAGVRSRPLGE---VRHALTHRLRVEVR 283  
 Db 253 EKRPAGIWMGLYCVPCF--ESLNGLSDFAKFSLTMADMDEQVATLHRLHRLLLTTPF 310

QY 284 GALW--ESEGDEPWKR-----LPKIME 304  
 Db 311 EAQMPSESPSDGIWIKPAHLKDYGLPKPLE 340

## RESULT 9

adenine glycosylase BH0931 [Imported] - *Bacillus halodurans* (strain C-125)  
 C:Species: *Bacillus halodurans*  
 C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: C83766  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a A:Reference number: AB3650; MUID:20512582; PMID:11058132  
 A:Accession: C83766  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-372 <STO>  
 A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BAB04650.1; GSPDB:G A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH0931  
 C:Superfamily: A/G-specific adenine glycosylase

Query Match 29.3%; Score 497.5; DB 2; Length 372;  
 Best Local Similarity 36.9%; Pred. No. 2.6e-31;  
 Matches 122; Conservative 54; Mismatches 120; Indels 35; Gaps 10;

QY 1 VEAMRKALLAMWYRENNAPLPWGEKDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLK 60  
 Db 10 ISTFONDLVTFWFSNHYRELPMRENKDPYRVWSEIMLQOTRVDIVIPYQAFMQPFTLE 69  
 QY 61 ALAAASLEEVLRVWOGAGYRRAEHLRLARSV-----ELPPSEAEELRGPLGCPYTA 115  
 Db 70 TLTAAREQVLAWEGLGYYSRRNLQSAVKEVESYGEVPSRKESIKLKGPGYTAG 129  
 QY 116 AVASIAFGERVAADVGNVRVLSR-LFARESPEKEKEFLAAGLLGEGV---DPGVWNOA 171  
 Db 130 AILSIAYDQPEPAVDGNVMWVLSRVLIIEEDIAKVKTRITFESLLYVLISKENSPFNOG 169  
 QY 172 LMELGATVCLPKPRRCGACPLGAFCRGKEAPGRYPAD---RKRRAKEERLVALVLGRKG 228  
 Db 190 LMEGLAVCLPTPTSGCLCPVRDHCRAFAVGVOEPLIKAKKKPKAKQLIAAVIRNEKG 249  
 QY 229 -VHLERL--EGRFGGLVGPPLFPPEELPGREAFGR---VRSR-----PLGE---VRHALT 274  
 Db 250 QVLIERRPEKLLAKMQFPVVELESTRKNAQOVLGDIYIHERFHLDAAVGEYVOTVEHVS 309  
 QY 275 HRLRVEVRGALWEGEGDEPWKRPLPKLMEK 305  
 Db 310 HLINIRYEXATYVG-----VPSLNDK 331

## RESULT 10

purine nucleosidase (EC 3.2.2.1) muty [Imported] - *Lactococcus lactis* subsp. *lactis* ( N:Alternate names: A/G-specific adenine glycosylase  
 C:Species: *Lactococcus lactis* subsp. *lactis*  
 C>Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malmme, K.; Weltsenbach, J.; Eh Genome Res. 11, 731-753, 2001  
 A:Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* A:Reference number: AB6625; MUID:21235166; PMID:11337471  
 A:Accession: AB6730  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-387 <STO>  
 A:Cross-references: GB:AE005176; PID:g12723767; PIDN:AAK04939.1; GSPDB:GN00146 A:Experimental source: strain IL1403  
 C:Genetics:  
 A:Gene: muty



C:Genetics:  
A:Gene: mult  
C:Superfamily: A/G-specific adenine glycosylase  
C:Keywords: DNA repair; iron-sulfur protein

[illegible]

**RESULT 14**

A/G-specific adenine glycosylase homolog [lnl1797] (imported) - Listeria innocua (strain AD1657)

A/G-specific adenine glycosylase  
C:Species: Listeria innocua  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AD1657  
R:Glaser, P.; Fraenkel, I.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, R.; Glaser, P.; Fraenkel, I.; Duchaud, E.; Durand, T.; Dussurget, O.; Entian, K.D.; Eschl, H.  
D.: Dominguez-Bernal, G.; Karslt, U.  
A:Sequence 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M.  
Ok, C.; Schleuter, T.; Simoes, N.; Thieriez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend  
A:Title: Comparative genomics of Listeria species  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AD1657  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-365 <G>AA>  
A:Cross-references: GB:AL592022; PIDN:CAC97028.1; PID:J16414300; GSPDB:GN00178  
A:Experimental source: strain ClIP11262  
C:Genetics:  
A:Gene: lnl1797  
C:Superfamily: A/G-specific adenine glycosylase

Query Match            28.4%; Score 483; DB 2; Length 365;  
Best Local Similarity 31.0%; Pred. No. 3,4e-30;  
Matches 107; Conservative 121; Indels 52; Gaps 10;

QY        1 VEAMRKALAWYRENAAPLPMPKGEKDPIRYLVSEVLLOQRVQALPYRRFLERPEPTLK 60  
   : :::::|::| | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db        14 ITAFQGLAVSYEAKMRKLIPWRENTEPIYRWISLIMQKQKVTFIYFREFMTQTPME 73  
QY        61 ALAASLEVLRVWGAGYRRRAHLRLARSV-----ELPSFAELRGILPGITYTA 115  
   : ::::|::| | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db        74 SFVAADADLTLMKEGLGYTSVRNLQIFAMQVMTDSGEVPSDLTTLLSKVGPTNG 133  
QY        116 AVASIVFGERRVAADVGVNRVYSRLPAKE---SPKRELALAQGLLPGCVDPGVNQ 171  
   : ::::|::| | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db        134 AITSIAYNQAEPAVDGNVMRVIAARVLETSEDIMKSTRTKEFEVLVQLIDQENPAAENG 193

RESULT 15  
D83320  
A/G-specific adenine glycosylase VC0452 [imported] - Vibrio cholerae (strain N16961 s  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: D83230  
R:Heldberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.  
Chadson, D.; Ermolaeva, M.D.; Yamathavan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers  
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483. 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: AB2035; NCID:20406833; PMID:10952301  
A:Accession: D82320  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-353 <HEIDI>  
A:Cross-references: GB:AE004132; GB:AE003852; NID:9654871; PIDN:AAE93625.1; GSPDB:CN  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
C:Gene: VC0452  
A:Map position: 1  
C:Superfamily: A/G-specific adenine glycosylase

[illegible]

Search completed: September 15, 2003, 13:53:45  
Job time : 41 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 19, 2005, 23:55:40 : Search time 111 Seconds  
(without alignment)  
6243.416 Million cell updates/sec

Title: US-09-938-901A-1  
Perfect score: 975  
Sequence: 1 atggagcgctgcgggaagc.....tagtccctcccgagcga 975

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/prodata/1/ina/5A COMB.seq:\*

2: /cgn2\_6/prodata/1/ina/5A COMB.seq:\*

3: /cgn2\_6/prodata/1/ina/5A COMB.seq:\*

4: /cgn2\_6/prodata/1/ina/5A COMB.seq:\*

5: /cgn2\_6/prodata/1/ina/5A COMB.seq:\*

6: /cgn2\_6/prodata/1/ina/5A COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	125.4	12.9	963	US-09-252-991A-6832	Sequence 6832, Ap
2	125.4	12.9	1401	US-09-252-991A-6886	Sequence 6886, Ap
3	125.4	12.9	1413	US-09-252-991A-6739	Sequence 6739, Ap
4	113	11.6	1551	US-09-489-039A-4788	Sequence 4788, Ap
5	107.2	11.0	549	US-09-252-991A-6783	Sequence 6783, Ap
6	104	10.7	1858	US-08-813-574-1	Sequence 1, Appl
7	87.2	8.9	2293	US-09-651-656-38	Sequence 38, Appl
8	87.2	8.9	2293	US-09-650-855-38	Sequence 38, Appl
9	86.6	8.9	4403765	US-09-103-840A-2	Sequence 2, Appl
10	86.6	8.9	4411529	US-09-103-840A-1	Sequence 1, Appl
11	82	8.4	1176	US-09-583-110-1922	Sequence 1922, Ap
12	82	8.4	9909	US-08-961-527-12	Sequence 12, Appl
13	81.2	8.3	968	US-09-864-866-46	Sequence 46, Appl
14	63.6	6.5	1212	US-09-107-532A-3654	Sequence 3654, Ap
15	62.4	6.4	1188	US-09-134-000C-1710	Sequence 1710, Ap
16	62.2	6.4	1227	US-09-385-028-23	Sequence 23, Appl
17	62.2	6.4	1227	US-09-726-614-23	Sequence 23, Appl
18	62.2	6.4	1227	US-09-385-028-23	Sequence 23, Appl
19	62.2	6.4	15079	US-09-385-040-23	Sequence 1, Appl
20	62.2	6.4	15079	US-09-385-040-23	Sequence 1, Appl
21	62.2	6.4	15120	US-09-385-040-1	Sequence 1, Appl
22	61.8	6.3	10444	US-09-543-681A-1143	Sequence 1143, Appl
23	61.4	6.3	1230025	US-09-198-452A-1	Sequence 1, Appl
24	59.6	6.1	20235	US-07-642-734C-3	Sequence 3, Appl
25	59.6	6.1	20235	US-08-439-009A-3	Sequence 3, Appl
26	58.8	6.0	1041	US-09-328-352-856	Sequence 856, Appl
27	56.6	5.8	2889	US-08-537-002A-4	Sequence 4, Appl

28	56.6	5.8	2889	US-08-863-010-4	Sequence 4, Appl
29	56.6	5.8	2889	US-09-024-429-4	Sequence 4, Appl
30	56.6	5.8	3600	US-08-537-002A-5	Sequence 5, Appl
31	56.6	5.8	3600	US-08-863-010-5	Sequence 5, Appl
32	56.6	5.8	3600	US-09-024-429-5	Sequence 5, Appl
33	55.6	5.7	2681	US-08-928-213B-7	Sequence 7, Appl
34	55.6	5.7	2363	US-09-818-780-22	Sequence 22, Appl
35	55.6	5.7	53500	US-09-266-965-76	Sequence 76, Appl
36	55.2	5.7	2634	US-08-941-936-1	Sequence 1, Appl
37	54.8	5.6	1404	US-09-252-991A-1725	Sequence 1725, Ap
38	54.8	5.6	2208	US-09-252-991A-2086	Sequence 2086, Ap
39	54.8	5.6	2586	US-09-252-991A-1804	Sequence 1804, Ap
40	54.8	5.6	3378	US-09-252-991A-2158	Sequence 2158, Ap
41	53.8	5.5	7560	US-09-754-112A-2	Sequence 2, Appl
42	53.8	5.5	12738	US-09-754-112A-1	Sequence 1, Appl
43	53.8	5.5	47981	US-09-679-279-1	Sequence 1, Appl
44	52.8	5.4	4257	US-08-690-473-1	Sequence 1, Appl
45	52.8	5.4	4257	US-09-259-821A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-09-252-991A-6832  
Sequence 6832, Application US/09252991A  
Patent No. 6531795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252, 991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074, 788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094, 190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6832  
LENGTH: 963  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6832

Query Match 12.9%; Score 125.4; DB 4; Length 963;  
Best Local Similarity 54.7%; Pred. No. 2.5e-17;  
Matches 311; Conservative 0; Mismatches 231; Indels 27; Gaps 2;

QY	55	CTCCCTGCGGCGGGAAGACCTTACCGGCTCTGTCGAGGTCCTTGCAG	114
DB	53	CTGCGCTGCGACAGGCGATCAACCCCTTACCGGTGTGGTCTCGAATCATGCTGCAG	112
QY	115	CAGACCCGGGTGAGAGGCGCTCCCTTATCCGCGCTTCTGAGGCGCTTCCGACC	174
DB	113	CAGACCCAGGTAGACACCTGCTCGTTACTTCTGACCGTTTATGACAGCTTGCCGAC	172
QY	175	CTGAAGGCGCTGCGCGCTTCTCTGGAAGAGTCTTAGGTCGAGGCGGCGG	234
DB	173	GTGAGGAGCATGCGCGCGCGCGCGCGAGGAGCAAGTCTGACCTGTGAGCGCGGCG	232
QY	235	TACTACCGGCGGCGGAGACCTTCAACCGCTGCGCGCGGAGGCT	279
DB	233	TACTACAGCGCGGCGGCGGAGACCTTCAACCGCGCGGAGGCT	277
QY	280	GAGGAGCTTCCCGCGAGCTTCCGCGGAGGCTTCTGAGTCTGAGGCGCTTACAC	339
DB	293	GAGGAGCTTCCCGCGAGCTTCCGCGGAGGCTTCTGAGTCTGAGGCGCTTACAC	352
QY	340	GCGGCGGCGGCTTCTGAGTCTTCTGAGGAGGCGGCTGAGCGGAGGCGG	399
DB	355	GCTGAGGCGATGCGGCTGCTGAGTGGGCTGCGGCGGCGGCGGATCTTCAACGCGG	412

QY 400 CGAGAGGCTCTCTCCCGGCTCTTGGCCCGGAAAGCCCAAGAGAGAGCTTTGGCC 459  
Db 413 AAGCGGTACTGCGCGCTACTGCGAGAGACGGCTATCCCGGCAACGAAAGTGGCC 472  
QY 460 CTGGCCAGAGGCTCTCTCTCCCGAGGCGGTGAGACCGGGGGGTGG-----AAC 507  
Db 473 AGGGGCTGTGGAAAGCGCGCAAGCTTCACTCCCGACGACGGGTCAACATACACC 532  
QY 508 CAGGCGCTCATGAGCTGGGGGCGAGGTCTGCTCTCGGAAAGCGCCCGCTGGGGGCC 567  
Db 533 CAGGCGATGATGACCTTGGGGCGACCTCTGACACGAGCAAGCCGAGTTGCTGCTT 592  
QY 568 TGGCCCGTGAAGGCGCTTCTGCGGGGGA 596  
Db 593 TGGCCGCTGTCTCGGCTCGCGCGCA 621

## RESULT 2

US-09-252-991A-6886/c  
Sequence 6886, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR APPLICATION NUMBER: 1999-02-18  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6886  
LENGTH: 1401  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6886

Query Match Best Local Similarity 12.9%; Score 125.4; DB 4; Length 1401;  
Matches 311; Conservative 0; Mismatches 231; Indels 27; Gaps 2;

QY 55 CTCCCTGGCGGGGAGAGAGACCTTACCGCGTCTGCTTCCAGAGTCTTTCGAG 114  
Db 1137 CTGGCTGGCAGACAGAGGATACCCCTTACCGGGTGTGGTCTCGAAATCATGCTGAG 1078  
QY 115 CAGACCGGGTGGAGACAGCCCTTCCCTATTACCGCGCTTTCGAGAGCGCTTCCACC 174  
Db 1077 CAGACCGAGTACAGACCGGTCTGCTTACCTTGAACCTTTCAATGACGCTTGCAGAC 1018  
QY 175 CTGAAGGCTGGCGCGGCTTCCCTGGAAGAGTCTTAAAGGTCTGGCAGAGGCGGAGC 234  
Db 1017 GTCCAGGAGACTGGCGGGGGCGGAGAGAGAGTCTTCACTGTGAGACCGGGCTCGGC 958  
QY 235 TACTACCGGGGGGGAACCTTCAACCGCTGGCCCGAAGCGT-----G 279  
Db 957 TACTACAGCGGTGGCGCAACCTGCAACAAGCCGCGCAATGTGTGTGAAGCGCATGCG 898  
QY 280 GAGGAGCTTCCCGAGCTTGGCGGAGCTTGGGGGCTTCTGTGCTGGGCTTTACACC 339  
Db 897 GGGGAGTTCCTCCCGACGTCGAGCACTGCGCAATGCGCGGATCGGCGCTCCACC 838  
QY 340 GCGGCGGAGTGGCTTCCCTTCCGAGAGCGGGTGGCGGGTGGAGCGGAAAGTCTC 399  
Db 837 GCTGAGAGCATGCGCAAGCTGTGATGGGCTGTGGCGGACGCAATCTCGAGCGGAAAGTCT 778  
QY 400 CCGAGGCTCTTCTCCGCTTTCGCGCGGAGAAAGCCCAAGAGAGAGCTTTTCGCC 459  
Db 777 AAGCGCTGATGAGCGCGCTTACCTGCGGAGAGAGAGCTATCCGCGCAACGAAAGTGGCC 718  
QY 460 CTGCGCCGAGGCTCTCTCCCGAGAGCGGTGAGCCCGGGGCTGTGG-----AAC 507

Db 717 AGGGCGTGTGGAAGCCCGCAAGCTTCAACCCCGACGACGGGTCAACACTACACC 658  
QY 508 CAGGCGCTCATGAGCTCGGGGCGACGGTCTGCGTCCGCAAGAGCCCGCTTGGGGGCC 567  
Db 657 CAGGCGATGATGACCTTGGGGCGACCTCTGACAGCGGAGCAAGCCCAATGCTGCTT 598  
QY 568 TGGCCCGTGAAGGCGCTTTCGCGGGGGA 596  
Db 597 TGGCCGCTGTCTCGGCTCGCGCGCA 569

## RESULT 3

US-09-252-991A-6739  
Sequence 6739, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR APPLICATION NUMBER: 1999-02-18  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6739  
LENGTH: 1413  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6739

Query Match Best Local Similarity 12.9%; Score 125.4; DB 4; Length 1413;  
Matches 311; Conservative 0; Mismatches 231; Indels 27; Gaps 2;

QY 55 CTCCCTGGCGGGGAGAGAGACCTTACCGCGTCTGCTTCCAGAGTCTTTCGAG 114  
Db 409 CTGGCTGGCAGACAGAGGATACCCCTTACCGGGTGTGGTCTCGAAATCATGCTGAG 468  
QY 115 CAGACCGGGTGGAGACAGCCCTTCCCTATTACCGCGCTTTCGAGAGCGCTTCCACC 174  
Db 469 CAGACCGAGTACAGACCGGTCTGCTTACCTTGAACCTTTCAATGACGCTTGCAGAC 528  
QY 175 CTGAAGGCTGGCGCGGCTTCCCTGGAAGAGTCTTAAAGGTCTGGCAGAGGCGGAGC 234  
Db 529 GTCCAGGAGACTGGCGGGGGCGGAGAGAGAGTCTTCACTGTGAGACCGGGCTCGGC 588  
QY 235 TACTACCGGGGGGGAACCTTCAACCGCTGGCCCGAAGCGT-----G 279  
Db 589 TACTACAGCGGTGGCGCAACCTTGCACAAAGCCGCGCAATGTGTGTGAAGCGCATGCG 648  
QY 280 GAGGAGCTTCCCGAGCTTGGCGAGCTTGGGGGCTTCTGTGCTGGGCTTTACACC 339  
Db 649 GGGGAGTTCCTCCCGACGTCGAGCACTGCGCAATGCGCGGATCGGCGCTCCACC 708  
QY 340 GCGGCGGAGTGGCTTCCCTTCCGAGAGCGGGTGGCGGGTGGAGCGGAAAGTCTC 399  
Db 709 GCTGAGAGCATGCGCAAGCTGTGATGGGCTGTGGCGGACGCAATCTCGAGCGGAAAGTCT 768  
QY 400 CCGAGGCTCTTCTCCGCTTTCGCGCGGAGAAAGCCCAAGAGAGAGCTTTTCGCC 459  
Db 769 AAGCGCTGATGAGCGCGCTTACCTGCGGAGAGAGCTATCCGCGCAACGAAAGTGGCC 828  
QY 460 CTGCGCCAGGCTCTCTCCCGAGAGCGGTGAGCCCGGGGCTGTGG-----AAC 507  
Db 829 AAGCGCTGTGGAGAGCGCGAAGCTTCAACCCCGACCAAGAGTCAACATACACC 888  
QY 508 CAGGCGCTCATGAGCTCGGGGCGACGGTCTGCTGCGGAAAGCGCCCGCTTGGGGGCC 567  
Db 889 CAGGCGATGATGACCTTGGGGCGACCTCTGACAGCGGAGCAAGCCCAATGCTGCTT 948

Qy 568 TGCCCCCTAGGGGCTTCTGCGGGGAA 596  
 Db 949 TGCCCGCTGTCTCCGGCTGCGCGGCA 977

# RESULT 4 US-09-489-039A-4788

; Sequence 4788, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709, 2004001

; CURRENT APPLICATION NUMBER: US/09/489, 039A

; PRIOR FILING DATE: 2000-01-27

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 4788

; LENGTH: 1551

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-4788

Query Match 11.6%; Score 113; DB 4; Length 1551;

Best Local Similarity 53.0%; Pred. No. 9.2e-15;

Matches 307; Conservative 0; Mismatches 245; Indels 27; Gaps 2;

Qy 53 CCTCTCCCTGCGGGGAGAAAGACCTTACCGCTCTGCTCTCGAGTCTCTTC 112  
 Db 560 CCTGCTGCGGAAATCGCCAGAGCCCTTACAAATATAGGCTCTCGAAGTATGTC 619  
 Qy 113 AGCAGACCGGGGTGAGCAGCGCTCTCTATTAACCGCGCTTCTGAGCGCTTCCA 172  
 Db 620 AACAAACCCAGGTGACACAGGTATCCCTATTGAAAGCTTTATGAGCGCTTCCCA 679  
 Qy 173 CCTGAAAGCCCTGCGCGCGGCTCTCTGAAAGGCTCTTAGGCTCTGAGCGGGGG 232  
 Db 680 CGGTGTGATCTTCGCCAAGCGCGCGCTGATGATGCTGATCTGTGACCGGCTG 739  
 Qy 233 GCTACTACCGCGGGGGAACAC-----CTGACCGCTGCGCGGAGCG 277  
 Db 740 GCTACTACCGCGGGGGAACACTTACATTAAGCGCGGAGCAAGTGGCAGACTGCA 799  
 Qy 278 TGGAGAGCTTCCCGAGCTTCCGAGCTTCCGAGCTTCCGAGCTTCCGAGCTTCA 337  
 Db 800 GCGGGGAATTCCTCCGAGCTTTCAGAAAGTGGCGGCTACCGCGCGCTCA 859  
 Qy 338 CGCGCGCGCGGCTGCGCTCATCGCTTGGGAGAGCGGCTGCGGCTGAGCGGAAG 397  
 Db 860 CGCGCGCGCGGCTTATCTCTTTCGCTGCTGATGATTCGATTCGAGCGGAAG 919  
 Qy 398 TCCGAGAGGCTCTCTCCGCTCTTTCGCGCGGAAAGCCCAAGAGAGAGCTTTG 457  
 Db 920 TGAAGCGCTGCTCGCGCGCTGCTATGCTGACGCGCTGCGCGGGAAGAGCTG 979  
 Qy 458 CCTGCGCCGAGGCTCTCTCCCGAGAGGCTGAGACCGGAGGCTGTG-----A 505  
 Db 980 AAAAAGCGCTGAGGACATCAGCAAGAGGTCAACCGCGAGAGGCGTGAAGCTTCA 1039  
 Qy 506 ACCAGGCTCTCATGAGCTCGGGGCAAGGCTGCTGCTGCGGAAAGCGCGCTTGGGG 565  
 Db 1040 ACCAGGCTATGATGATCTCGGGGCAATGTTTGAACCGGCTGAGAGCGGAAGTGGAGC 1099  
 Qy 566 CTGCGCCCTAGAGGCTTCTGCGCGGAGAGAGAGCC 604  
 Db 1100 TGTGTCTGAGCAAGCGCTGCTGCTTACGCAACC 1138

RESULT 5  
 US-09-252-991A-6783  
 ; Sequence 6783, Application US/09252991A

; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252, 991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074, 788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094, 190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 6783

; LENGTH: 549

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6783

Query Match 11.0%; Score 107.2; DB 4; Length 549;

Best Local Similarity 54.4%; Pred. No. 1.3e-13;

Matches 248; Conservative 0; Mismatches 193; Indels 15; Gaps 1;

Qy 55 CTCCCTGCGGGGAGAGACCTTACCGGCTCTGCTCTCGAGGCTCTTTCAG 114  
 Db 60 CTGCGCTGCGAGAGGAGATCAACCCCTACCGGCTGTGGGTCTCGAATCATGTCAG 119  
 Qy 115 CAGACCCGGGTGAGAGAGCCCTCCCTTATTAACCGCGCTTCTGAGCGCTTCCACC 174  
 Db 120 CAGACCCAGTCAAGCACCTGCTCGGTACTTCAACGTTTCAAGCAGCGTTCGAGC 179  
 Qy 175 CTGAAGGCTTGGCGCGGCTTCCCTGAAAGTCTTAAAGTCTGAGAGGGGGGGG 234  
 Db 180 GTGAGGCACTGCGCGGCGGCGGAGAGAGTCTCAACTGTGAGACCGGCTGCGG 239  
 Qy 235 TACTACCGCGGGGAGAACCTTCAACCGCTGCGCGGAGCT-----G 279  
 Db 240 TACTACAGCGCTGCGCGGCACTGCAAGAGCGGCAATGTGTGAGCGGAGTGG 299  
 Qy 280 GAGAGCTTCCCGAGCTTCCGCGGAGCTTCCGAGGCTTCTGAGTCTGAGGCTTCAACC 339  
 Db 300 GGGGAGTTCCTCCGCGAGCTGAGCAATCGCGGAATCCCGGCTATCGGCGCTCAC 359  
 Qy 340 GCGCGCGGCTGCTTCAATCGCTTCCGAGAGCGGCTGCGGCTGAGAGCGGAACCTC 399  
 Db 360 GCTGAGCCATGCGAGCTGCTGATGAGGCTGCGGAGCAGATCTTCAAGCGGACGTC 419  
 Qy 400 CGAAGGCTCTTCCCGCTTCCGCGGAGAGCGGAGAGAGAGAGAGCTTTGCGC 459  
 Db 420 AAGCGCTACTGCGCGCTTCACTGCGGAGAGAGCGGCTATCCCGGAGACCGAAGTGGC 479  
 Qy 460 CTGCGCCAGAGGCTCTCTCCGAGAGGCTGAGACCG 495  
 Db 480 AGGGGCTGTGGAGAGCGCGGAGAGCTTACCCG 515

# RESULT 6 US-08-813-574-1

; Sequence 1, Application US/08813574

; Patent No. 6013473

; GENERAL INFORMATION:

; APPLICANT: Wei, Ying-Fei

; TITLE OF INVENTION: Human Muty

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/813,574
FILING DATE:
CLASSIFICATION: 536
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 60/013,132
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: ATG50002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1858 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-813-574-1

```

```

Query Match      10.7%; Score 104; DB 3; Length 1858;
Best Local Similarity 54.0%; Pred. No. 6,7e-13;
Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;

```

```

69 GGAAGAGACCCCTTACCGGCTCTGATCTCCGAGGTCCTTTCGACGACACCCTGGTGA 128
129 GCAGGCTCTCCCTATTACCCGCTTTTTCGAGCCCTTCCACCTGAAGGCTTGGC 188
558 CACTGTGATCACTACTATACCGGATGATGCAAGATGGCTTACACTGACGAGACTTGGC 617
189 CGGGGCTTCCCTGGAAGAGGTCCTTAGGCTGCGAGGGGGGCTACTACCGGCGGC 248
618 CAGTCTCTCTGAGAGAGGTGAATCACTCTGGGCTGGCTGAGCTACTATTCTGTGG 677
249 GGAACACTCCACCGCTGCGCCGGAAGCTGAGAGGCTTCCCGAGGCTC----- 300
678 CCGGCGCTGACGAGAGGAGCTCGAAGGTGTGAGAGAGCTAGAGGGCCACATGCCAG 737
301 -----GCCAGCTTGGGGGCTTCTGTGCTCGGCTTACACCGGCGGCGGT 350
738 TACAGCAGAGACCTTGCAGAGCTCTGCTGGGCTGGGCGCTACACAGCTGGGCGCAT 797
351 GAGCTCCATCGCTTTCGGGAGCGGGTGGCGGCTGAGAGGAAAGTCCGAGGCTTCT 410
798 TGCTCTATCGCTTTCGCTGGCCAGGCAACCGGTGTGTGAGTGGCACTGACGAGGTGCT 857
411 CTCCGCTCTTTCGCTC-----GGGAAAGCCCAAGAGAGAGAGCTTTTCGCTGCCAG 468
858 GTGCGGTGTGCGAGCACTTGTGCTGATTCACAGAGACCTTTCCTCCAGAGCTCTG 917
469 GGCCTCTTCCCGAGAGGCTGAGCCG-----GGGTGTGGAACAGAGGCTCTCAT 518
918 GGGTCTAGCCGAGAGCTGTGTGAGCCAGCCCGGCAAGAAATTTCAACAGAGCAT 977
519 GAGAGCTGGGGGACAGGTCTGCTGCGGAACGCGGCTTTCGAGGGGCTTCCCTTAA 578
978 GAGAGCTAGGGGACAGGTGTGTGATCCACAGGCGCCACTGTGAGCAGCAGTCCCTGTGA 1037
579 GGCCTTTCGCGGG 592
1038 GAGCTGTGCGGG 1051

```

RESULT 7

```

US-09-651-656-38
Sequence 38, Application US/09651656
Patent No. 6340566
GENERAL INFORMATION:
APPLICANT: MCCLUTCHEN-MALONEY, SANDRA
APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE
POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,
TITLE OF INVENTION: DNA DAMAGE AND DNA MISMATCHES
FILE REFERENCE: IL-10689
CURRENT APPLICATION NUMBER: US/09/651,656
CURRENT FILING DATE: 2000-08-29
PRIORITY APPLICATION NUMBER: 60/192,764
PRIORITY FILING DATE: 2000-03-28
NUMBER OF SEQ ID NOS: 106
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38
LENGTH: 2293
TYPE: DNA
ORGANISM: Escherichia coli
US-09-651-656-38

```

```

Query Match      8.9%; Score 87.2; DB 3; Length 2293;
Best Local Similarity 54.1%; Pred. No. 1.9e-09;
Matches 210; Conservative 0; Mismatches 163; Indels 15; Gaps 1;

```

```

53 CCTCCCTGGCGGGGAGAGACCTTACCGGCTCTGCTCTCCGAGGTCCTTCTGC 112
1161 CTCTGCTCCGCAATTGACAGAGCCCTTACCAAGATGCTCTGAGAGTATGTTGC 1220
113 AGCAGACCCGGGTGAGAGAGCCCTTCCCTATTACCGCCCTTTCGAGCCCTTCCCA 172
1221 AACAACTCAGGTGCGACCTTATCCCTTATTGAAAGCTTTATAGCGCGCTCCCA 1280
173 CCTGAGAGCCCTGCGCGGCTTCCCTGGAAGGTCCTTAGGTCCTGCGAGGGGCGG 232
1281 CGGTATCGATCTGCGCAATGCGCGCTCGAGAGGTTCTCACTGTGAGACGGGCTTG 1340
233 GCTACTACGGGGGCGGAGCACTCCACCGCTGGCCCGAAGCTGAG----- 282
1341 GCTATTACCCCGCGCGCAATCTGCAATTAAGCGGCAACAAAGTGGCGCACTTACACG 1400
283 -----GAGCTTCCCGAGCTTCCGCAAGCTTTCGGGGCTTCTGCTGCGGCTTACA 337
1401 GCGGTAAATTCGCGAATCTTTGAGAAATGACACTCCCGGCTGCGGCTTCCA 1460
338 CCGCGGCGGCTGAGCTTCATCGCTTTCGGGAGCGGAGCGGCGGCTGAGCGGAAAG 397
1461 CCGAGGCGGATTCCTGCTTCTCTGGGTAAAGCACTTCCGATTCGACGGTAAAG 1520
398 TCGGAGAGCTCTTCCGCTCTTTCG 425
1521 TCAAGCGCTGCGGCGGCTGTATGC 1548

```

```

RESULT 8
US-09-650-855-38
Sequence 38, Application US/09650855
Patent No. 6365355
GENERAL INFORMATION:
APPLICANT: MCCLUTCHEN-MALONEY, SANDRA
APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA
MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA
TITLE OF INVENTION: MISMATCHES
FILE REFERENCE: IL-10284
CURRENT APPLICATION NUMBER: US/09/650,855
CURRENT FILING DATE: 2000-08-29
PRIORITY APPLICATION NUMBER: 60/192,764
PRIORITY FILING DATE: 2000-03-28
NUMBER OF SEQ ID NOS: 106
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38

```

LENGTH: 2293  
 TYPE: DNA  
 ORGANISM: Escherichia coli  
 US-09-650-855-38

Query Match 8.9%; Score 87.2; DB 3; Length 2293;  
 Best Local Similarity 54.1%; Pred. No. 1.9e-09;  
 Matches 210; Conservative 0; Mismatches 163; Indels 15; Gaps 1;

53 CCTCTCCCTGGCGGGGAGAGAACCTTACCGCGCTCTGCTCCGAGCTCTTCTGC 112  
 1161 CTCTCCCTGGCGAAATGCAAGACGCTTACAAATATGCTCTCAAGATGATGTC 1220  
 113 AGCAGACCGGGTGGAGACGCGCTTATTAACCGCGCTTCTGAGCGCTTCCCA 172  
 1221 AACAACTCAAGTTGCGACCGCTTATTCCTATTTGAACGCTTTATGCGCGCTTCCA 1280  
 173 CCTGAAAGCCCTGGCGCGCTTCCCTGGAAGAGTCTTGAAGTCTGCGAGCGCGG 232  
 1281 CGGTACCGATCTCGGCATATGCGCGCTGAGAGATTCATCTTGTGACCGGCTTG 1340  
 233 GCTACTACCGCGGGGAGAACCTTCCACCGCTTGGCCGGAAGCTGAG----- 282  
 1341 GCTATTACGCGCGCGCGCAATGCAATAAGCGGCACAAGTGGCGACTTACAG 1400  
 283 -----GAGCTTCCCGGAGCTTCCGCGAGCTTGGGGGCTTCTGTCTCGGCGCTTACA 337  
 1401 GCGGTAATTCCTCGGAACCTTTGAGAGATTGACGACATCGCGCGCTTCCA 1460  
 338 CGCGCGCGCGGTGCGCTTCATCGCTTGGGAGCGGTGGCGCGGTGAGCGGAGACG 397  
 1461 CGGAGCGCGCATCTCTCTGCTTCTCTGGTAAAGACTTCCGATTCTCGACGTTACG 1520  
 398 TCCGAGGCTCTCTCCGCTCTTTCG 425  
 1521 TCAACGCGTCTGGCGCGCTGTATGC 1548

## RESULT 9

US-09-103-840A-2  
 Sequence 2, Application US/09103840A  
 Patent No. 6294328  
 GENERAL INFORMATION:  
 APPLICANT: FLEISCHMAN, Robert D.  
 APPLICANT: WHITE, Owen R.  
 APPLICANT: FRASER, Claire M.  
 APPLICANT: VENTER, John C.  
 TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 FILE REFERENCE: 24366-2007.00  
 CURRENT APPLICATION NUMBER: US/09/103,840A  
 CURRENT FILING DATE: 1998-06-24  
 NUMBER OF SEQ ID NOS: 2  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 2  
 LENGTH: 4403765  
 TYPE: DNA  
 ORGANISM: Mycobacterium tuberculosis  
 FEATURE:  
 OTHER INFORMATION: CDC 1551  
 OTHER INFORMATION: "n" bases at various positions throughout the sequence  
 OTHER INFORMATION: represent a, t, c or g  
 US-09-103-840A-2

Query Match 8.9%; Score 86.6; DB 3; Length 4403765;  
 Best Local Similarity 50.2%; Pred. No. 6.3e-09;  
 Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

22 CTCCTCGCCTGTATCCGGAAGAACCGCGCTTCCCTGCGCGGGG---GAGAAAGAC 78  
 4022631 CTTCTCGCTTGTATCAGGATCGACCGGATCTGCTCGCGAGAGCCCGGTGTCAAC 4022690  
 79 CCTTACCGGTCTGTCTCGAGGATCTTCTGACAGACCGCGGTGAGAGAGCGCTTC 138

4022691 CCGTGGCAGATCTGTGTACGAGATTACGTGACAGACACGCGCGCGCGGTGCTG 4022750  
 139 CCTTATTACCGCGCTTCTTGTGAGGCTTTCCACCTGTAAGGCCCTGGCGGCTTCC 198  
 4022751 GCGATCTGCGCGACCTGGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGG 4022810  
 199 CTGGAAGAGTCTTAAAGGCTTGGCAAGGGGGGGGCTATACCGGAGGGGAGAACCTC 258  
 4022811 ACCCGGATGTATTACGCGCTTGGGCAAGCTGGGCTTATCCAGAGAGCTCAAGCTTAA 4022870  
 259 CACCGCTGGCGCGGAAGCTG-----GAGAGCTTCCCGGAGCTTGGCC 303  
 4022871 CACAGGTGGCCACCGGTATGCGCGCGGACCAACATGAGTGTGTGCGACATATCAG 4022930  
 304 GAGCTTGGGGGCTTCTGTGTCTGGGCTTACACCGCGCGCGGTGCTTCATGCGC 363  
 4022931 ATCTGTGTACCTTCCGCGCGCTGGGAGCTACCGCGCGCGGTGGCTGTTCGCT 4022990  
 364 TTGCGGAGAGGCTGGCGCGGTGGAGCGGAACTCCGAGGGTCTCTCCGCGCTT- 422  
 4022991 TACCGCAGCGGTGTGGGTGGACACCAATGTGGCGGTGTGGCGCGCTT 4023050  
 423 -----GCGCGGAAAGCCCAAGAGAAAGAGCTTTCGCTTCCGCGCAGGCG 471  
 4023051 CACGCGCGCGCGAGCGCGGTGCGGTGCGGTGCGCGCGGACACCGCGCTTGGCG 4023110  
 472 CTCCTCCCGAGGCGGTGACCCGCGGAGTGTGAACCAAGCCCTCATGAGCTCGGGCC 531  
 4023111 CTGTTCGCGCACCGGAGACGCGCGCTGTAATTTGGTGTGCTGTATGAGTTGGGTGG 4023170  
 532 AGGTCTGTGCTGCGCAAGACGCGCTTGGGGGGCTGCGCTTGAAGGCTTCTGCGCG 591  
 4023171 ACGGTGTGACCGCGCGCACACCGCGGTGCTTATGCGCTGTGATGCTGCGCATG 4023230  
 592 GGAAGAGAGCGCGCGCGCTACCCGCGCGCAGAGACGCGCGGAGAGAGAGCGC 651  
 4023231 CGGCATCCCGGTATTCGCGGTGAGACGCTC---GCGCGCGCGGAGCGCTTACAC 4023287  
 652 CTCGTGCGCTTCTCTCTCTCGGGCGG 678  
 4023288 GGAACCGACCGCAAGTCCGCGAGCG 4023314

## RESULT 10

US-09-103-840A-1  
 Sequence 1, Application US/09103840A  
 Patent No. 6294328  
 GENERAL INFORMATION:  
 APPLICANT: FLEISCHMAN, Robert D.  
 APPLICANT: WHITE, Owen R.  
 APPLICANT: FRASER, Claire M.  
 APPLICANT: VENTER, John C.  
 TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 FILE REFERENCE: 24366-2007.00  
 CURRENT APPLICATION NUMBER: US/09/103,840A  
 CURRENT FILING DATE: 1998-06-24  
 NUMBER OF SEQ ID NOS: 2  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 1  
 LENGTH: 4411529  
 TYPE: DNA  
 ORGANISM: Mycobacterium tuberculosis  
 OTHER INFORMATION: H37Rv  
 US-09-103-840A-1

Query Match 8.9%; Score 86.6; DB 3; Length 4411529;  
 Best Local Similarity 50.2%; Pred. No. 6.3e-09;  
 Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

22 CTCCTCGCCTGTATCCGGAAGAACCGCGCTTCCCTGCGCGGGG---GAGAAAGAC 78  
 CTTCTCGCTTGTATCAGGATCGACCGGATCTGCTCGCGAGAGCCCGGTGTCAAC 4022690



Query Match 8.4%; Score 82; DB 4; Length 9909;  
 Best Local Similarity 52.2%; Pred. No. 2,7e-08;  
 Matches 218; Conservative 0; Mismatches 185; Indels 15; Gaps 1;

QY 22 CTCCTGCTGTGATACCGGAAACGCGCCCTCCCTGCGGGGAGAGACCTT 81  
 DB 5808 CTCCTGCTGTGATACCGGAAACGCGCCCTCCCTGCGGGGAGAGACCTT 5749  
 QY 82 TACCCGCTGTGATACCGGAAACGCGCCCTCCCTGCGGGGAGAGACCTT 141  
 DB 5748 TATCAATGTGATACCGGAAACGCGCCCTCCCTGCGGGGAGAGACCTT 5689  
 QY 142 TATTAACCGCTGTGATACCGGAAACGCGCCCTCCCTGCGGGGAGAGACCTT 201  
 DB 5688 TACTAGAAAGATTTTGGACGTGTTCCACGTGTGAAAGTGTGCACTGGCCCTGAG 5629  
 QY 202 GAAGAGTCTTAAAGGTCTGCGAGGGGGCGGCTA-----CTACCGGCG 246  
 DB 5628 GAGAGTTTACTAAAGCTTGGAGGGGCTTGGGCTATTATTCTGAGTTGCAATATGCA 5569  
 QY 247 GGGGAACCTCCACCGCTGCGCCGAGCGTGAAGACCTTCCCGAGCTTCCCGAG 306  
 DB 5568 GCTGAGCCGACAGATTATGACTGACTTGTGGCCATTCCAAATACCTATGAAGA 5509  
 QY 307 CTTCGGGGCTTCTGTCTCGGGCTTACACCGCGCGCGGTGGCTTCAATGCGCTTC 366  
 DB 5508 ATTTCAGCTTAAAGGATTTGACCTTACAGCGAGGACATTTCAGATTTGCTTTT 5449  
 QY 367 GGGAGCGGGTGGCGCGGTGACGCGGAACGTTCGGAGGCTCTTCCGCTCTTTC 424  
 DB 5448 AACTTGCTGAGCAGACTGTAGTGTATATGTCATCGGGGTCTTGGCGGCTGTTTG 5391

RESULT 13  
 US-09-864-866-46  
 ; Sequence 46, Application US/09864866  
 ; Patent No. 6723548  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lloyd, R. Stephen  
 ; APPLICANT: McCullough, Amanda K.  
 ; APPLICANT: Nguyen, Khia  
 ; TITLE OF INVENTION: DNA REPAIR POLYPEPTIDES AND METHODS OF USE  
 ; FILE REFERENCE: 265, 00170101  
 ; CURRENT APPLICATION NUMBER: US/09/864,866  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/206,279  
 ; PRIOR FILING DATE: 2000-05-23  
 ; NUMBER OF SEQ ID NOS: 49  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 46  
 ; LENGTH: 968  
 ; TYPE: DNA  
 ; ORGANISM: Micrococcus luteus  
 US-09-864-866-46

Query Match 8.3%; Score 81.2; DB 4; Length 968;  
 Best Local Similarity 47.2%; Pred. No. 3e-08;  
 Matches 300; Conservative 0; Mismatches 318; Indels 18; Gaps 1;

QY 44 AGCCCGCCCTCCCTGCGGGGAGAGAACCTTACCGCGCTGTGCTCCGAG 103  
 DB 200 ACCCGTACCGCTTCCCGAGCTTGAAGACCGCTTCAAGCTGTGTCGACCG 259  
 QY 104 TCTTCTGAGAGACCGCGGTGAGACGAGCCCTCCCTATTATACGCGCTTTCGAGC 163  
 DB 260 TGTCTGCGCCGACGACGAGCGTGGCGGTCAAGCAGCAGCGCGGCGTTCGCC 319  
 QY 164 GCTTTCACCGCTGAGAGCGCTTGGCGGCTTCCCTGGAAGAGTCTTGGGTTCG 223  
 DB 320 GCTTTCGAGATCCGAGATGCGCGGCGCAGCGAGCGCTGCGAGAGCTGTGC 379  
 QY 224 AGGGGCGGGCTACTACCGG-----CGGGGGAACACTTCACCGGC 265

DB 380 GCTCCAGGGGTTCTACCGGAACAGGCTCCGCGATCTCGGCTGTGCCAGAGCTCG 439  
 QY 266 TGGCCCGAAGCGTGAAGAGCTTCCCGAGCTTGGCCGAGCTTCCGGGGCTTCTGTGTC 325  
 DB 440 TGGCCCGGACGACGCGGAGGTCCCGCCGCTTCTGAGGACCTTGTGGGCTTCCCGGG 439  
 QY 326 TGGGCTTACACCGCGGGGCGGTGCTTCCATCGCTTGGGAGAGCGGTGCGCGCG 385  
 DB 500 TGGGCGGACAGCGGCTTGTGTGTCTGCGCAAGCTTCCGCGAGCCCGGATACCG 559  
 QY 386 TGGACGGAAGCTCCGAGAGGCTCTCTCCGCTTCTTCCCGGGAAGCCCCAGAGA 445  
 DB 560 TGGACAGGACTTCCGCGCGCTTCCCGCGCTTCCGAGGTTACAGAGACGAGACCGCG 619  
 QY 446 AGAGCTTTCGCTTCCGCGCGCTTCTCCCGAGGCGTGAACCGGGGGTGTGA 505  
 DB 620 GTAAAGTGAACACGCCGTGGCGGCTGTTCCTCCCGGAGCTGACATCTCT 679  
 QY 506 ACCAGGCTCTCATGAGCTCGGGGCGCAGGCTGCTGCGGAAACGCGCCGTTGCGGG 565  
 DB 680 CCGACGCGTGAATCTTCCAGCGCGCGCGCTGTGCTGACGCGCGCGCGCGCTGCGGG 739  
 QY 566 CCGCCCGCTAGAGGCTTCTGCGGGGAGAGAGCGCCCGGCGCTACCCCGCGCCA 625  
 DB 740 GGTGCCGATCGCCCGCTGTGCTGCTCTACGCGCGGGGAGACGACCCCGAGCGGG 799  
 QY 626 GGAAGCGCGGCGGAGAGAGCGGCTGTGCGCC 661  
 DB 800 CGCGGCGCTGTGCTGCTGAGCTCAAGCCGCGCC 835

RESULT 14  
 US-09-107-532A-3654  
 ; Sequence 3654, Application US/09107532A  
 ; Patent No. 6583275  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 7310  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 ; STREET: 100 Beaver Street  
 ; CITY: Waltham  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02354  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: CD-ROM ISO9660  
 ; COMPUTER: PC  
 ; OPERATING SYSTEM: <Unknown>  
 ; SOFTWARE: ASCII  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/107,532A  
 ; FILING DATE: 30-Jun-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/085,598  
 ; FILING DATE: 14 May 1998  
 ; APPLICATION NUMBER: 60/051571  
 ; FILING DATE: July 2, 1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ariniello, Pamela Deneke  
 ; REGISTRATION NUMBER: 40,489  
 ; REFERENCE/DOCKET NUMBER: GTC-012  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (781) 893-5007  
 ; TELEFAX: (781) 893-8277  
 ; INFORMATION FOR SEQ ID NO: 3654:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1212 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: circular



```

; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-1710

Query Match      6 4%;      Score 62.4;  DB 4;  Length 1188;
Best Local Similarity 48.8%;  Pred. No. 0.00023;
Matches 211;  Conservative 0;  Mismatches 206;  Indels 15;  Gaps 1

OY      8  CCTGGCGGAAAGCCCTCTCTGCTGTACCGGAAACGCCGCCCTCTCTCTGCGG 67
        | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      53  CATTCACGAAATTTTATGCTTGAGTGAACGAGAAACGAAATTACCTTGGCGAG 112
        | | | | | | | | | | | | | | | | | | | | | | | | | |
        53  CCGGACGACCTTACCGGCTCTGTGCTCCGAGCTCTTCTGACACAGACCGGGTGG 127
        | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Query Match	Best Local Similarity	Score	DB	Length
Matches 211; Conservative	48.8%;	62.4;	4;	1188;
		Pred. No. 0.00023;		
		Mismatches 206;	Indels 15;	Gaps 1
QY	8	CCGCGGGAAGCCCTCCTCGCTGTACCGGAAAACGCCCGGCCCTCCCTCGCGG	67	
DB	53	CATTCGAGAAAGTTTTTTAGCTGTATGACGAGAAAACGCAATTTCCTTGGGAG	112	
QY	68	GGGAGAGACCTTACCGCGTCCGAGTCTTCTGACAGACCCGGTGG	127	
DB	113	CGAATACGATGCAATCGATTGGATTTCGAAATATGCTACCAACAACTGCGTAG	172	
		CGAATACGATGCAATCGATTGGATTTCGAAATATGCTACCAACAACTGCGTAG	187	

Matches	211;	Conservative	97	100%
QY	8	CCTGGGGGAAACCCCTCTCGCCCTGGACCGGGGAAACGCCGCCCCCTCTCCCTGGCGGG	67	
Db	53	CATTCAGAGAAATTTTTCATCTGGTATGACGAGAAACGCAATTACTTGGCGAG	112	
QY	68	GGGAGAGAGACCCTTACCGCGCTCGCTCCGAGTCTCTCGACGAGACCCGGGTG	127	
Db	113	CGATTACAGATGATTCGATTTTGGATTCTGAAATTAATGCTACACAAATCCGGTGA	172	
QY	128	AGCAGGCCCTCCCTTACCGCGCTTTCGAGCCGCTTCCACCTGAAAGGCCCTGG	187	
Db	173	ATAAGTCATTGATTTTATTCATTCGATTTATGAAATGATTTCCGACGATTCAAGATTGA	232	
QY	188	CGCGGGCTTCCCTGGAAGAGTCTTGAAGTCTGGACGGGGCGGGCTACTACGGCGGG	247	
Db	233	CGGAAGCCCGATGATGATTTGTGAAAGCTTGGGAAGGTTATAGTTACTATTCACGAG	292	
QY	248	C-----GGAACACCTTCCACCGCGCTTCCGCCGAAGGTGGAGAGACTTCCC	292	
Db	293	CGCGTATTTTAAAGTGGAGCCCAACAATTTGTTTCAGAAATTTGGTGGAAATGCTG	352	
QY	293	CGAGCTTGCAGACTTCGGGGGCTTCGTGCTCGGGCTTACACCGCGCGCGGTGG	412	
Db	353	ACGACATTCAMAAATTTTGGAGTTTAAAGAGATCGGCCCTTATCGGCTGTGCGATTG	412	
QY	353	CTTCATCTCCCTTCGGGGAGCGGGTGGCGGCGGTGGACGGAACTCTCGAGAGGTCTCT	412	
Db	413	GCAGTATTCGCTTTATCTCCGCAACCACTAATGACGTTAACGTCATGGGGGTAGTGA	472	
QY	413	CCGCGCTCTTGC	424	
Db	473	GCGGTATTTTG	484	



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2005, 04:17:41 ; Search time 643 Seconds  
(without alignments)  
8712.661 Million cell updates/sec

Title: US-09-938-901A-1

Perfect score: 975  
Sequence: 1 atgagggcctgcggcgaaagc.....tagtcccccctccgcagcga 975

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
20: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
21: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	974	99.9	975	US-09-938-901-1	Sequence 1, Appli
2	112.4	11.5	939	US-10-156-761-4691	Sequence 4691, Ap
3	112.4	11.5	9025608	US-10-156-761-1	Sequence 1, Appli
4	104	10.7	1854	US-09-925-301-484	Sequence 484, App
5	104	10.7	1878	US-09-925-301-484	Sequence 484, App
6	92.8	9.5	528	US-09-974-300-1535	Sequence 1583, Ap
7	91.2	9.4	5850	US-10-437-963-1325	Sequence 1325, Ap
8	89.2	9.1	1660	US-10-425-114-32132	Sequence 32132, A
9	89.2	9.1	1749	US-10-425-115-92673	Sequence 92673, A
10	86.6	8.9	912	US-10-216-817-31	Sequence 31, Appl
11	86.6	8.9	1312	US-10-216-817-21	Sequence 21, Appl
12	86.2	8.8	1425	US-10-437-963-35788	Sequence 35788, A

C	13	82	8.4	9909	8	US-08-961-527-12	Sequence 12, Appl
C	14	82	8.4	9909	16	US-10-158-844-12	Sequence 12, Appl
C	15	81.2	8.3	968	9	US-09-864-866-46	Sequence 46, Appl
C	16	72.6	7.4	2256646	17	US-10-470-565-1	Sequence 1, Appli
C	17	71.4	7.3	930	15	US-10-156-761-4577	Sequence 4577, Ap
C	18	71.4	7.3	9025608	15	US-10-156-761-1	Sequence 1588, Ap
C	19	68.2	7.0	991	16	US-10-398-221-1588	Sequence 3596, Ap
C	20	68.2	7.0	3155	16	US-10-398-221-3596	Sequence 3629, Ap
C	21	66.8	6.9	11058	15	US-10-156-761-3629	Sequence 42, Appl
C	22	66.8	6.9	18876	15	US-10-328-079-42	Sequence 34, Appl
C	23	66.8	6.9	61944	15	US-10-328-079-34	Sequence 56678, A
C	24	66.4	6.8	953	18	US-10-425-115-56678	Sequence 5759, Ap
C	25	66.4	6.7	4667	18	US-10-723-860-5759	Sequence 5404, Ap
C	26	64.6	6.6	1330	15	US-10-156-761-5404	Sequence 3, Appli
C	27	63.8	6.5	8324	15	US-10-402-842-3	Sequence 1, Appli
C	28	63.8	6.5	47988	15	US-10-402-842-1	Sequence 23116, A
C	29	63.6	6.5	2428	17	US-10-437-963-23116	Sequence 17, Appl
C	30	62.8	6.4	2731748	17	US-10-297-465A-1	Sequence 23, Appl
C	31	62.4	6.4	32768	9	US-09-070-927A-17	Sequence 17, Appl
C	32	62.2	6.4	1327	15	US-10-458-201-23	Sequence 13, Appl
C	33	62.2	6.4	15120	15	US-10-458-201-1	Sequence 1, Appli
C	34	61.8	6.3	1949	17	US-10-437-963-50882	Sequence 50882, A
C	35	61.4	6.3	1230025	16	US-10-289-762-1	Sequence 1, Appli
C	36	61.2	6.3	975	10	US-09-938-901-1	Sequence 8597, Ap
C	37	60.6	6.2	1041	17	US-10-437-963-8597	Sequence 12606, A
C	38	60.6	6.2	1484	16	US-10-424-599-105239	Sequence 5, Appl
C	39	60.4	6.2	566	17	US-10-437-963-12606	Sequence 19, Appl
C	40	60.4	6.2	2036	17	US-09-938-901-5	Sequence 1, Appli
C	41	59.8	6.1	1029	10	US-10-205-032-19	Sequence 9047, Ap
C	42	59.8	6.1	5355	15	US-10-205-032-1	Sequence 4667, Ap
C	43	59.4	6.1	60196	15	US-09-864-408A-9041	
C	44	59.4	6.1	272	11	US-10-156-761-4667	
C	45	58.8	6.0	1122	15	US-10-156-761-4667	

## ALIGNMENTS

RESULT 1  
US-09-938-901-1  
; Sequence 1, Application US/09938901  
; Publication No. US20030008291A1  
; GENERAL INFORMATION:  
; APPLICANT: Kuramitsu Seiki,  
; TITLE OF INVENTION: GENE ENCODING DNA REPAIR ENZYME  
; FILE REFERENCE: PH-1261-US  
; CURRENT APPLICATION NUMBER: US/09/938,901  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: JP2001-47762  
; PRIOR FILING DATE: 2001-02-23  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 975  
; TYPE: DNA  
; ORGANISM: Thermus thermophilus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(975)  
US-09-938-901-1

Query Match 99.9%; Score 974; DB 10; Length 975;  
Best Local Similarity 100.0%; Pred. No. 3.9e-220;  
Matches 974; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	2	TCGAGGCTGCGGGAAGCCCTCTCGCTGTACCGGAAACGCCGCCCTCCCT	61
DB	2	TGAGGCTGCGGGAAGCCCTCTCGCTGTACCGGAAACGCCGCCCTCCCT	61
QY	62	GCGCGGGGGAAGAACCTTACCGCGTCTGTCTCCAGAGTCTTTCGACAGACCC	121
DB	62	GCGCGGGGGAAGAACCTTACCGCGTCTGTCTCCAGAGTCTTTCGACAGACCC	121

```
QY 122 GGGTGGAGAGGCGCTCCCTATTAACCGCCCTTTCTGAGGCGCTTCCCAACCTGAAAG 181
Db 122 GGGTGGAGAGGCGCTCCCTATTAACCGCCCTTTCTGAGGCGCTTCCCAACCTGAAAG 181
QY 182 CCTTGGCGCGGCTTCTCCCTGAAAGAGTCTTAAGGCTGAGGCGGCGGCTTAAGC 241
Db 182 CCTTGGCGCGGCTTCTCCCTGAAAGAGTCTTAAGGCTGAGGCGGCGGCTTAAGC 241
QY 242 GCGCGGCGGAGACACCTCCACCGCTGCGGCCGAAAGCTGAGAGACTTCCCGAGCTTCC 301
Db 242 GCGCGGCGGAGACACCTCCACCGCTGCGGCCGAAAGCTGAGAGACTTCCCGAGCTTCC 301
QY 302 CCGAGCTTCCGAGGCTTCTGCTGCTGAGGCTTACACCGGCGGCGGCGGCTTCCATCG 361
Db 302 CCGAGCTTCCGAGGCTTCTGCTGCTGAGGCTTACACCGGCGGCGGCGGCTTCCATCG 361
QY 362 CTTTCCGAGGAGCGGCTTCCGCGGCTGAGCGGAAAGCTCCGAGAGGCTCTCTCCGCTCT 421
Db 362 CTTTCCGAGGAGCGGCTTCCGCGGCTGAGCGGAAAGCTCCGAGAGGCTCTCTCCGCTCT 421
QY 422 TCGCCCGGAAAGCCCAAGAGAGAGAGCTTTTCGCGCTCCGCGGCGGCTCTCCGCTCT 481
Db 422 TCGCCCGGAAAGCCCAAGAGAGAGAGCTTTTCGCGCTCCGCGGCGGCTCTCCGCTCT 481
QY 482 AGGCGCTGAGACCGGCGGCTTGTGAACAGAGCTTCAATGAGCTCGGAGCAAGCTCTGCG 541
Db 482 AGGCGCTGAGACCGGCGGCTTGTGAACAGAGCTTCAATGAGCTCGGAGCAAGCTCTGCG 541
QY 542 TCGCGAAAGCGGCGGCTTGTGAGGCGGCTTCCCTGAGGCGCTTTCGCGGCGGAGAGG 601
Db 542 TCGCGAAAGCGGCGGCTTGTGAGGCGGCTTCCCTGAGGCGCTTTCGCGGCGGAGAGG 601
QY 602 CCGCGGCGGCTTACCGCGGCGGCGGAGAGCGCGGCGGAGAGAGCGCTCTGCGCGC 661
Db 602 CCGCGGCGGCTTACCGCGGCGGCGGAGAGCGCGGCGGAGAGAGCGCTCTGCGCGC 661
QY 662 TCGCTCTCTCGGCGGAGAGGAGGCTGCACTTGAAGAGCTTGAAGGCGCTTCAAGGCGC 721
Db 662 TCGCTCTCTCGGCGGAGAGGAGGCTGCACTTGAAGAGCTTGAAGGCGCTTCAAGGCGC 721
QY 722 TCTAAGGAGCGGCTTCTTCCCTGAGAGCTTCCGCGGCGGAGAGCGGCTTCCGCGC 781
Db 722 TCTAAGGAGCGGCTTCTTCCCTGAGAGCTTCCGCGGCGGAGAGCGGCTTCCGCGC 781
QY 782 TGAAGTCTAAGCGCTTGAAGAGGCTGCGCAAGCGCTTCAACCGAGGCTTCCGCTG 841
Db 782 TGAAGTCTAAGCGCTTGAAGAGGCTGCGCAAGCGCTTCAACCGAGGCTTCCGCTG 841
QY 842 AGGTGCGGCGGCGCTTGTGGAAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
Db 842 AGGTGCGGCGGCGCTTGTGGAAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
QY 902 AGCTCATGAGAGAGTCTCCGAGAGGCGCTTCCCTCTGCTCATGAGGCGGCTTCCG 961
Db 902 AGCTCATGAGAGAGTCTCCGAGAGGCGCTTCCCTCTGCTCATGAGGCGGCTTCCG 961
QY 962 CCTTCCGAGAGCA 975
Db 962 CCTTCCGAGAGCA 975
```

## RESULT 2

```
US-10-156-761-4691
/ Sequence 4691, Application US/10156761
/ Publication No. US20030119018A1
/ GENERAL INFORMATION:
/ APPLICANT: OMURA, SATOSHI
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: ISHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
```

```
/ APPLICANT: HATTORI, MASAHIRA
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ PRIOR FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 4691
/ LENGTH: 939
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(939)
US-10-156-761-4691
```

```
Query Match 11.5%; Score 112.4; DB 15; Length 939;
Best Local Similarity 52.1%; Pred. No. 2,6e-17;
Matches 369; Conservative 0; Mismatches 306; Indels 33; Gaps 4;
```

```
QY 4 GAGGCTTGGCGGAAAGCCCTCTCGCTGATACCGGAGAAAGCGCCCGCTTCCCTG 63
Db 73 GAGGCTTGGCGGAAAGCCCTCTCGCTGATACCGGAGAAAGCGCCCGCTTCCCTG 63
QY 64 CGGAG---GGAAGAGACCTTTACCGGCTGCTGATCCGAGTCTTTCGAGAGAGC 120
Db 133 CGGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 192
QY 121 CGGATGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
Db 193 CGGATGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 252
QY 181 GCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
Db 253 GAGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 312
QY 241 CGGCGGCGGAGACCTTCAACCGGCTGAGC-----GGAAGCTGAGAGAG 285
Db 313 CGGCGGCGGCTGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 372
QY 286 CTTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 345
Db 373 GTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 432
QY 346 GCGGTGCTCTCATCGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405
Db 433 GCGGTGCTCTCATCGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 492
QY 406 GTCTCTCCGCGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 459
Db 493 GTCTCTCCGCGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 552
QY 460 -----CTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 510
Db 553 CGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 612
QY 511 GCGCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 570
Db 613 GCGCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 672
QY 571 CCGCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 630
Db 673 CCGATGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 732
QY 631 CGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 678
Db 733 CGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
```

## RESULT 3

```

US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018a1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, TUNO
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

Query Match	11.5#	Score	112.4	DB	15	Length	9025608
Best Local Similarity	52.1#	Pred.	No.	6e-18			
Matches	369	Conservative	0	Mismatches	306	Indels	33
							Gaps
							4
Qy	4	GAGGCTTGCGGGAAAGCCCTCTCGCTGTATCCGGGAAACCGCCGCCCTCCCTCG	63				
Dp	5736491	GAGGCTCTGCACGCGCCAGGTATCGCATGGTTTCAGCAGACGACGCGCTGACCTCCGTGG	5736432				
Qy	64	CGGGG---GGAAGAGAACCTTACCGCGCTCTGGTCTTCCAGGTCCTTCTGCAGACACC	120				
Dp	5736431	CGCGGCCCAAGCGCGGCGCCCTGGGGGGGTGATGGTCAAGTGGTTCATGTTGCACACAGCA	5736372				
Qy	121	CGGGTGGAGCAGGCGCTCCCTCTATTACCGCGCTTTCGAGCGCTTTCACACCTGAAG	180				
Dp	5736371	CGGTCACACCGCGTGTGCTGCGCTTACAGAGAGATGGCTCGCCGCTGGCCCGCCCGCC	5736312				
Qy	181	GCCCTGGCCGCGGCTTCCCTGGAAGAGTCTTAAGGTCTTGACAGGAGGCGGCTACTAC	240				
Dp	5736311	GACCTTCCCAAGGAGGCGCGCCGCGGAAGCCGTCCGCGCTGGGCGCCCTTCGCTTACCC	5736252				
Qy	241	CGGGGGGGGAGCACTTCCACCGGCTGGCC-----CGAAGGTGGAAGAG	285				
Dp	5736251	CGGGGCGCGCTGCGCCCTGCACAGGCGCGCGCTCGCCATTAACGGAACGCAACAGCGGAC	5736192				
Qy	286	CTTTCGCCGAGCTTTCGCGGAGCTTTCGGGAGCTTCTGGTCTCGGAGCTTACACCGCGAG	345				
Dp	5736191	GTAACCAACGAGACGCGCCAGCTGTGCGGTGCGCGGATTCGGCGAATACACGCGCGCG	5736132				
Qy	346	GCGGTGCGCTTCATCGCTTTCGGGAGCGGCTGGCGCGGTGAGCGGAACTTTCGAGAG	405				
Dp	5736131	GCGGTGCGCTGTTCGCTTACGGGCAACGCAACCGCTTTCGACACCAACGTCGCGCGT	5736072				
Qy	406	GTCCTTCGCGCGCTTTCGCGCGGGGAAAGCCCAAGAGGAGGAGGCTTTTGCC-----	459				
Dp	5736071	GTCCTTCGCGCGCGCTTCACCGGTGTGCAATACCGCGCCCAACGCGCACACCGACCGAG	5736012				
Qy	460	-----CTGCGCCAGGAGCTTCTTCGCCAGGAGCGTGAACCCGAGGAG--TGTGAACAG	510				
Dp	5736011	CGCAAGCTCGCGCGCGGTGTCTCCCGAGGACGAGAGCACCGGTCCCGCTGGGCGCC	5735952				
Qy	511	GCCTCATGAGAGCTTCGGGGGCAACGCTCTGCTCGCGAAGACGAGCCCGTGTGGGGAGCTTCG	570				
Dp	5735951	GCGTCATGAGAGCTTCGGGCGCGGTGTCTGCACGCGGAAAGACGAGACTGTGCACCGCTGC	5735892				

```

OY      571      CCCCAGGGGCGCTTCTGCGCGGGGGAAGAGAGCGCCCCGGGCGGTACCCCGCGCCAGAGAG 630
DB      5735891  CCGATCGCGCGCACTGCGCCTTGGCGGACTCTGGCGGGCAAGCCGAGACACGACGGGCGCGCG 5735892
OY      631      CGCCGGGCGAAGAGAGAGCGCCTCTGTGCGCCTCTGTCCTCTCTCGGCGCG 678
DB      5735891  CGCGCGCGCGACGCTACGCGCGGTACGACCGCTCAGGTCTCGCGCGAGG 5735784

RESULT 4
US-10-469-992-9
; Sequence 9, Application US/10469992
; Publication No. US20040096863A1
; GENERAL INFORMATION:
; APPLICANT: Liveoh, Zvi
; APPLICANT: Paz-Eilzsur, Tamara
; APPLICANT: Blumenstein, Sara
; TITLE OF INVENTION: METHODS AND KITS FOR DETERMINING A RISK TO DEVELOP CANCER, FOR EVALUATING THE EFFECTIVENESS AND DOSAGE OF CANCER THERAPY AND FOR CORRELATING THE TITLE OF INVENTION: ACTIVITY OF A DNA REPAIR ENZYME AND A CANCER
; FILE REFERENCE: 26315
; CURRENT APPLICATION NUMBER: US/10/469,992
; CURRENT FILING DATE: 2003-09-16
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 1854
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-469-992-9

```

```

Query Match 10.7%; Score 104; DB 16; Length 1854;
Best Local Similarity 54.0%; Pred. No. 2,2e-15;
Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;

QY GGAGAGAGACCTTACCCGCTCTGCTCCAGGCTCTTCTGCAGACAGACCCGGTGA 128
Db |||||
QY GGAGAGAGAGGCGATATGCTGTGTGGTCTCAAGGTCATGTCAGACAGCCAGGTTGC 553
Db |||||
QY GCAGGCCCTCCCTATTACCGCGCTTCTGAGGCTTCCACCTGAAGCCTGGC 188
Db |||||
QY CACTGTGATCAACTACTATACCGATGATGATGAGAGATGGCTTACACTCAGAGACTGGC 613
Db |||||
QY CGCGGCTTCCCTGGAAGAGTCTTAGGGTCTGAGAGGGGGGGTACTACCGGGGGC 248
Db |||||
QY CAGTGTCTTCCCTGAGAGAGTGAATCAACTCTGGGCTGGCTTACTATTCTGTGG 673
Db |||||
QY GGAACACCTCCACCGGCTGGCCGAGCGTGAAGAGCTTCCCGCAGCTTC----- 300
Db |||||
QY CCGGCGGCTGCAGAGAGGAGCTCGAAGTGTGTAGAGAGCTTAGGGGCGACATGCGACG 723
Db |||||
QY 301 -----GCCGAGCTTCCGGGGGCTTCTGTCTCGGGCTTACACCGCGGCGGT 350
Db |||||
QY 734 TACAGCAGAGACCCCTGCAGACAGCTCTGCTGGCGGCGCTCAACAGCTGGGCGCAT 793
Db |||||
QY 351 GGCCCTCATGCGCTTCCGGGAGCGGGATGGCGCGGTGAGCGGAACTCGCGAGGGTCT 410
Db |||||
QY 794 TGCCCTCTATCGCTTTGGCCAGAGCAACCGGTGTGGATGGCAAGTGCACAGGGTCT 853
Db |||||
QY 411 CTCGCCGCTCTTGGCCC--GGGAAAGCCCAAGAGAAAGACTTTTCGCCCTCGCCAG 468
Db |||||
QY 854 GTGCCGATTCGAGAGCATTTGGTGTGATCCAGAGCAACCTGTTTCCAGCAGCTTG 913
Db |||||
QY 469 GGCTCTCTCCCGAGGGCGTGAACCG-----GGGTTGTGAATCAAGGCTTCAT 518
Db |||||
QY 914 GGGTCTAGCCAGAGCTGTGGAACCCAGCCGAGCAGAGATTTCAACCAAGCAAGCAT 973
Db |||||
QY 519 GGAGCTCGGGGCAACGCTGCTGCTGCCGAACGCGCCGTTGCGGGGCTGGCCCCCTAG 578
Db |||||
QY 974 GGAAGTAAAGGAGCAAGTGTATCCCAAGCGCCCACTGTGCAAGCAAGTCCCTGTGA 1033
Db |||||
QY 579 GGCTCTTGCAGG 592

```

Db 1034 GAGCCTGTGCGGG 1047

## RESULT 5

US-09-925-301-484  
Sequence 484, Application US/09925301  
Patent No. US20020052308A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA106  
CURRENT APPLICATION NUMBER: US/09/925,301  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05882  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1694  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 484  
LENGTH: 1878  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1446)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-301-484

Query Match 10.7%; Score 104; DB 9; Length 1878;  
Best Local Similarity 54.0%; Pred. No. 2,2e-15;  
Matches 299; Conservative 0; Mismatches 225; Indels 30; Gaps 3;

QY 69 GGNAGAGACCTTACCGGCTCTGCTCCAGGCTCTTCCGACGACACCGGAGTGA 128  
DB 509 GGNAGAGCGGCGATATGCTGTGTGCTCAGAGGTCATGCTGCGACGACCGAGTTC 568  
QY 129 GCGAGCGCTCCCTATTACCGCGCTTTCGAGCGCTTCCACCTGAAAGGCTTGGC 188  
DB 569 CACTGTGATACATCTACTATACCGGATGAGTGCAGAAAGGCTTACCTGACGAGACTTGGC 628  
QY 189 GCGGCGCTTCCCTGAGAGGCTCTTGAAGGCTTGGCGAGGCGGCGGCTACTACCGGCGGAC 248  
DB 629 CAGTGCTTCCCTGAGAGGCTGATCAACTTGGCGCTGGCTGCTATTTCTGCTGG 688  
QY 249 GGAACACTTCCACCGCGCTGCGGCGGAGCTGAGAGGCTTCCCGAGCTTC----- 300  
DB 689 CCGGCGGCTGAGAGGAGGAGTCCGAAAGTGTATGAGAGGCTAGGCGGCGACATGCCACG 748  
QY 301 -----GCCGAGCTTGGGGGCTTCTGTGCTCTGGGCTTACACCGCGGCGGCTG 350  
DB 749 TACAGCAGAGACCTGCGACGAGCTCTGCTGGCGGCGGCTACACAGCTGGGGCT 808  
QY 351 GGCTCTTACCTGCTTGGGAGCGGCTGCGGCGGAGGAGCGTCCGAGGCTCT 410  
DB 809 TGGCTCTTATCGCTTGGCGAGCAGCGGCTGTGTGATGAGCACTAGCAGGCTGCT 868  
QY 411 CTCGCGCTCTTGGCC--GGGAAAGCCCAAGAGAGGCTTTCGCTGCGCTGCGAG 468  
DB 869 GTGCGGTGTCCGAGCATTTGGTGTGATCCAGGACCTTGTTCCTCCAGAGCTCTG 928  
QY 469 GGCTCTCTCCCGAGGCGGTGAGCCG-----GGGCTGTGAACCAAGGCTCTCAT 518  
DB 929 GGGCTTACGCCAGCAGCTGTGTGATCCAGCCGCGGCGAGAGATTTCACCAAGGCGCAT 988  
QY 519 GAGGCTCGGAGCAGCGCTGTGCTGCGGAAAGCGCCCGCTTGGGCGCTGCGCTTACG 578  
DB 989 GAGGCTAGGAGGCGCAGTGTGTATCCCAAGCGGCCACTGTGACGACGAGTGTGTA 1048  
QY 579 GGCCTTCTGCGGCG 592  
DB 1049 GAGCCTGTGCGGG 1062

## RESULT 6

US-09-974-300-1583  
Sequence 1583, Application US/09974300  
Patent No. US20020146721A1  
GENERAL INFORMATION:  
APPLICANT: Berka, Randy M.  
TITLE OF INVENTION: Methods for Monitoring Multiple Gene  
FILE REFERENCE: 10085,500-US  
CURRENT APPLICATION NUMBER: US/09/974,300  
CURRENT FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: 09/680,598  
PRIOR FILING DATE: 2000-10-06  
PRIOR APPLICATION NUMBER: 60/279,526  
PRIOR FILING DATE: 2001-03-27  
NUMBER OF SEQ ID NOS: 8481  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1583  
LENGTH: 528  
TYPE: DNA  
ORGANISM: Bacillus licheniformis  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)...(528)  
OTHER INFORMATION: n = A,T,C or G  
US-09-974-300-1583

Query Match 9.5%; Score 92.8; DB 9; Length 528;  
Best Local Similarity 53.4%; Pred. No. 1,2e-12;  
Matches 227; Conservative 0; Mismatches 183; Indels 15; Gaps 1;

QY 17 AAGCCTCTGCGCTGTGATCCGGAAGACCGCGGCTTCCCTGCGGCGGAGAGG 76  
DB 43 AAGATTATTTCTGTGATGAGCAGAGAAACGAGTCTGCTTGGCGCTGACAGCG 102  
QY 77 ACCCTTACCGGCTCTGTGCTCCGAGGCTTTCGACAGACCGGCTGAGACGCTGA 162  
DB 103 ACCCTTATAGGTGTGGTGTGGAGTATGCTTACGACAGACGAGTGTGACGCTGA 162  
QY 137 TCCCTTATACCGCGCTTTCGAGCGCTTTCACACCTTGAAGCCCTGGCGGCTT 196  
DB 163 TTCTTACTTCAACAATTATCGAAAGTTTCCGACGCTGAAAGCTGCTGAACCG 222  
QY 197 CCCTGAAAGGCTCTTGAAGTCTGCAAGGCGGCGGCTACTTACCGCGGCGGAACAC 256  
DB 223 ACAGAGAAAGTGTAAAGCTGGAAGACCTGGATATCTATCAAGGCTTGGAACT 282  
QY 257 TCCACCGC-----CTGGCCGAGCGTGAAGAGCTTCCCGAGCTTGC 301  
DB 283 TGCAGAGCGCTGTCAAGGAAGTTCAAGAACGATGAGGCTGCTCTTCAAG 342  
QY 302 CCGAGCTTGGGAGCTTCTGTGCTTGGGCTTACACCGCGCGGCTGCTTCATCG 361  
DB 343 AGGAATTCCGAGCTTAAAGGCTGCGCTTATCAAGAGGCGGCTGAGCACTCG 402  
QY 362 CCTTGGGAGCGGCTGCGCGGCTGAGCGGAACTCCGAGGCTCTTCCCGCTCT 421  
DB 403 CTTACATCAGCGGCTCCCGCGGTAGACGGAATGTCAATGGGCTCATGTCAGGATTTC 462  
QY 422 TCGCC 426  
DB 463 TTTC 467

## RESULT 7

US-10-437-963-1325  
Sequence 1325, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437, 963  
NUMBER OF SEQ ID NOS: 204966  
SEQUENCE OF SEQ ID NOS: 204966  
LENGTH: 5850  
TYPE: DNA  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_101202C.1  
US-10-437-963-1325

Query Match 9.4%; Score 91.2; DB 17; Length 5850;  
Best Local Similarity 54.4%; Pred. No. 1.9e-12;  
Matches 211; Conservative 0; Mismatches 168; Indels 9; Gaps 1;

65 GGGGGGAGAGACCTTACCGGCTGCTCCGAGGTCCTTCGACAGACCCGG 124  
263 GCGAGAGAGAGAGGCGTACGCGGTGTGGTTCGAGGTGATGTCGACATACCGCG 322  
125 TGGACAGGCGCTCCCTTATACCGCGCTTCTGAGCGCTTCCACCTGAAAGCC 184  
323 TGCCGCTGTGTGCTGACTACTCTCCGCTGATGAGCGCGCGCCACCTGAGAGCC 382  
185 TGGCGCGGCTTCTCTGAGAGGTCTTAAAGGTCTGCAAGGCGCGGCTACTACGCG 244  
383 TCGCGCTGCGACGAGAGAGGTGACGAGATGTGGCGGCGCTCGGCTATTACGCGA 442  
245 GGGGGGAGACCTTACCGGCTGCGCGCGCGAGCGT-----GAGAGAGCTTCCCGCA 295  
443 GGGCTCGATTCTTCTCGAGGAGCAAGCAATTTGMAAAAGCGAGTTCCTCGCA 502  
296 GCTTCCGAGCGCTTCCGCGCTTCTGCTGCTGCGGCTTACACCGCGCGCGGTGCGCT 355  
503 CAGCATCAGCACTCCGTGAAGTTGCGGATGCGGATTTACACAGAGAGCATTTGCTT 562  
356 CCATCGCTTCTGCGGAGCGGCTGCGCGCGGTGACGAGAACTCCGAGGCTTCTTCC 415  
563 CCATGCGCTTCAATGAGGTGTCTGTTGTGATGGAATGCGAGTTATCAGAGGC 622  
416 GCCTTCTGCGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 443  
623 TTACGCTATTCTCTGATTAACCAAGAG 650

RESULT 8  
US-10-425-114-32132

Sequence 32132, Application US/10425114  
Publication No. US2004003488A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/425, 114  
NUMBER OF SEQ ID NOS: 73128  
SEQUENCE OF SEQ ID NOS: 73128  
LENGTH: 1660

TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: UC-ZMFLB73265D07\_FLI  
US-10-425-114-32132

Query Match 9.1%; Score 89.2; DB 16; Length 1660;  
Best Local Similarity 50.7%; Pred. No. 6.9e-12;  
Matches 282; Conservative 0; Mismatches 253; Indels 21; Gaps 2;

60 CTGGCGGGGAGAGAGACCTTACCGCGCTGCTCCGAGGTCCTTCGAGAGAC 119  
278 CGGAGCGAGAGAGAGAGCGGTACCGGCTGCGGTGTCCGAGGTATCTGACAGAC 337  
120 CCGGCTGAGAGAGAGCGCTTCCCTATTACCGCGCTTCTGAGCGCTTCCACCTGAA 179  
338 GCGGGTCCCGTGTCTGTCTTACTACAGCGGTGATGCGCGGTGCGCGCGCTACG 397  
180 GCGCGTGGCGCGCGCTTCCCTGAGAGAGGTCTTACGCTGAGCGAGGGCGGCTACTA 239  
398 AAGCTTCCCGCGCGCTTACGAGAGAGGTGAAAGATGTGGCGGCTTGGCTACTA 457  
240 CCGCGCGCGGAGACCTTCC-----ACCGCTGCGCGCGAGAGCGTGAAGAGCTTCC 290  
458 CCGTAGGCTCATTTCTGCTGAGAGAGCAAGCATCATCAAAAGGGGTGTTTC 517  
291 CCGAGCTTCCCGAGCTTCCGCGGCTTCTGCTGCTGCGGCTTACACCGCGCGCGT 350  
518 TTGCAAGCGGTAAAGCTTCTGAGAGGTGTGAGCATTTGGGATTTACAGCTGAGACAA 577  
351 GCGCTTCATCGCTTCCGAGAGAGCGGTGCGCGGTGAGAGGAGAGCTCCGAGGCTCT 410  
578 CCGTTCAATGCTTCTTCAAGAGAGGTGTCCAGTTGTGATGAAATGTATACAGATCAT 637  
411 CTCGCGCTTCTTCCCGGAGAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 458  
638 CAGCAGGCTTTACCATTTGTGACACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 697  
459 CCTCGCGAGCGCTTCTTCCCGAGAGCGGTGAGAGCGGAGGTGAGAGAGAGAGAGAG 518  
698 GAGCTCTGAGGTCAATGTGTGACCTTTGAGAGCAGAGAGAGAGAGAGAGAGAGAG 757  
519 GAGAGCTCGGAGAGAGAGGTGCTGCTGCGAGAGAGCGCGGTGCGCGCGCGCGCGCTAG 578  
758 GAGAGTAG 817  
579 GCGCTTCTGCGCGGAG 594  
818 TAGCAGCTGCCAAGCG 833

RESULT 9  
US-10-425-115-92673

Sequence 92673, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/425, 115  
NUMBER OF SEQ ID NOS: 369326  
SEQUENCE OF SEQ ID NOS: 369326  
LENGTH: 1749  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_184509C.1  
US-10-425-115-92673

Query Match 9.1%; Score 89.2; DB 18; Length 1749;  
 Best Local Similarity 50.7%; Pred. No. 6.9e-12;  
 Matches 282; Conservative 0; Mismatches 253; Indels 21; Gaps 2;

```

QY 60 CTGGCGGGGGAAGAACCCCTTACCGGCTCTCGAGTCTCTTCTGACAGAC 119
DB 278 CGGACGAGAGAGAGAGAGGCGTACCGGCTGTGGGTCTCGAGTGTATGTCACAGAC 337
QY 120 CCGGCTGAGACAGGCTCTCTTATTAACCGGCTTCTGAGCGCTTCCACCGTAA 179
DB 338 GCGGGTCCCGTGGTCTCGCTTACTACGAGCGGTGATGGCGGGTGGCGACGCTAG 397
QY 180 GGGCTTGGCGGCTTCTCTGAGAGGCTTAAAGGCTGTGACAGGGGCGGCTACTA 239
DB 398 AAGCTGCGCGCGCTACGAGGAGGAGTGAACGAGATGTGGCGGGTCTTGCTACTA 457
QY 240 CCGGCGGGGAGACCTCC-----ACCGCTGGCGCGGAGCGGTGAGAGCTTCC 290
DB 458 CCGTAGGGCTGATTTCTGTGAGGAGGAGCAAGCATCATCGAAGGGGTGTGTTCC 517
QY 291 CCCGAGCTTCCCGAGCTTGGGGGCTTCTGTGTCTGGGCTTTACACCGCGCGGCT 350
DB 518 TTGACACGCGCTTACCGCTTGTGAGGTCTGTGATTTGCGGATTTACACAGCTGAGCAAT 577
QY 351 GGGCTTCATCGCTTGGGAGCGGCTGCGCGGTGAGCGGAACTCCGAGGCTCT 410
DB 578 CGCTTCAATTGCTTCAACGAGGTTGTCCAGTTGTGATGGAATGTGATACAGATCAT 637
QY 411 CTCGCGCTTTCGCGCGGAGAAAGCCCAAGAGAA-----GGAGCTTTTCG 458
DB 638 CAGCAGGCTTTACACCATTTGTCGACCAACCAAGGATCTTCAACAGTGAAGATCTG 697
QY 459 CCGCGCCAGGCGCTCTCCCGAGGCGGTGACCCGCGGGGTGTGAAACAGGCGCTCAT 518
DB 698 GACCTCTGAGTCAATATGTGTTGACCTTTGAGACAGAGACTTCAACCAAGAGATGAT 757
QY 519 GAGAGCTCGGGGACACGCTGTGCTGCGCCGAAAGCGGCGGCGGCGCTAGG 578
DB 758 GAGGTAGAGAGCACTTATGTAGCAAGCAAGGCTGTGCTCTCAATGCCACGCTC 817
QY 579 GGGCTTTCGCGGGGG 594
DB 818 TAGCAGCTGCAAGCG 833

```

## RESULT 10

US-10-216-817-31  
 ; Sequence 31, Application US/10216817  
 ; Publication No. US20030129619A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GICOUEL, BRIGITTE  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
 ; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
 ; TITLE OF INVENTION: IN GENES OF THE MULTIFAMILY  
 ; FILE REFERENCE: 03495.0233-00000  
 ; CURRENT APPLICATION NUMBER: US/10/216,817  
 ; CURRENT FILING DATE: 2002-11-13  
 ; PRIOR APPLICATION NUMBER: 60/311,824  
 ; PRIOR FILING DATE: 2001-08-14  
 ; PRIOR APPLICATION NUMBER: 60/313,523  
 ; PRIOR FILING DATE: 2001-08-21  
 ; NUMBER OF SEQ ID NOS: 32  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 31  
 ; LENGTH: 912  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 US-10-216-817-31

Query Match 8.9%; Score 86.6; DB 15; Length 912;  
 Best Local Similarity 50.2%; Pred. No. 3.1e-11;  
 Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

```

QY 22 CTCCTGCTGCTGATCCGGGAAACGCCCGCCCTCCCTGCGGGG---GAGAGAGAC 78
DB 61 CTCTGCTGCTGCTGATAGGATGCGACCGGATCTGCTTGGCGAGACCCGCTGTACAGC 120
QY 79 CTTACCGGCTCTGCTGCTGAGGCTCTTCTGACAGACAGCCGGGTGAGACAGGCGCTC 138
DB 121 CCGTGGCAATCTGTGTCAGCGAGTTCACTGTGACAGACAGCCCGCGCGCGGCTGCTG 180
QY 139 CCTATTACCGCGCTTCTTGTGAGCGCTTCCACCGCTGAAAGCCCTGCGCGGCTTCC 198
DB 181 GCGATGTGCGCGGACGTGGGTGCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCGCG 240
QY 199 CTGGAAGAGTCTTGAAGGTCTGCGAGGCGGCTACTACCGCGGCGGAGAACCTTC 258
DB 241 ACCGCGCATGTGTTATCGCGCTGGGCAAGCTGAGCTATCCAGCGAGACCAAGGCTTA 300
QY 259 CACGCGCTGCGCGGAGCGTG-----GAGAGCTTCCCGGAGCTTCCG 303
DB 301 CACGAGTGGCGCACCGTATGCGCGCGACACATGATGTGCTCCAGATATCGAG 360
QY 304 GAGCTTGGGGGCTTCTGCTGCTGAGCTTACACCGCGCGGCGGCTGCTCATCGCC 363
DB 361 ATCTGTGACCTTGGCGCGGCTGCGGAGCTACACCGCGCGGCTGCTGCTGCT 420
QY 364 TTGCGGAGCGGCTGCGCGGCTGAGACGGAAGCTCCGAGGCTCTTCCGCTCTT- 422
DB 421 TACGCGCAGCGGCTGCGGCTGAGACCAAGTGTGCGGCTGTGCTGCTGCTGCT 480
QY 423 -----CGCGCGGAGAAAGCCCAAGAGAGAGCTTTCGCTGCGCGCAGGCG 471
DB 481 CAGGCGCGCGCGAGCGGCTGCGCGCATCGGTGCGCGCACGCGCGAGCTTGTGGC 540
QY 472 CTCCTCCCGAGGCGGTGAGACCCGCGGCTGTGAAACAGAGCCCTCATGAGCTCGGGCC 531
DB 541 CTGTGCGCGACCGCGAGACCGCGCTGAAATTTTGGCTGCGCTGATGAGATGGGTGCG 600
QY 532 AGGCTGCTGCTGCGGAAACGCGCGGCTGCGGCGCTGCGGCTGAGAGGCTTTCGCGG 591
DB 601 ACGGTGTGACCGCGCGCACACCGCGGCTGCGGCTGAGCTGCGCTGAGCTGCGCATGG 660
QY 592 GGAAGAGAGCGCGCGCGGCTGACCGCGCGCCAGAGAACCGCGCGGAGAGAGAGCGC 651
DB 661 CCGCATGCGCGGTATTCGCGCGTGGACGCTC---GCGCGCGCGGCGGAGCTTACAC 717
QY 652 CTCGTGCGCTGCTGCTCTCGGGCGG 678
DB 718 GGAACGAGCGCGCAAGTCCGCGGACGG 744

```

## RESULT 11

US-10-216-817-21  
 ; Sequence 21, Application US/10216817  
 ; Publication No. US20030129619A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GICOUEL, BRIGITTE  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING MULTIDRUG  
 ; TITLE OF INVENTION: RESISTANT STRAINS OF M. TUBERCULOSIS HAVING MUTATIONS  
 ; TITLE OF INVENTION: IN GENES OF THE MULTIFAMILY  
 ; FILE REFERENCE: 03495.0233-00000  
 ; CURRENT APPLICATION NUMBER: US/10/216,817  
 ; CURRENT FILING DATE: 2002-11-13  
 ; PRIOR APPLICATION NUMBER: 60/311,824  
 ; PRIOR FILING DATE: 2001-08-14  
 ; PRIOR APPLICATION NUMBER: 60/313,523  
 ; PRIOR FILING DATE: 2001-08-21  
 ; NUMBER OF SEQ ID NOS: 32  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 21  
 ; LENGTH: 1312  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 US-10-216-817-21

Query Match 8.9%; Score 86.6; DB 15; Length 1312;  
 Best Local Similarity 50.2%; Pred. No. 2.9e-11;  
 Matches 345; Conservative 0; Mismatches 309; Indels 33; Gaps 4;

```

QY 22 CTCCTCGCTGTGTACCGGAAAAACGCCCGCCCTTCCCTGGCGGGG---GAGAAGAC 78
DB 261 CTTCTCGCTGTGTATCAGGATCGCAACCGGATCTCCCTGGGAGAGCCCGGTGCAGC 320
QY 79 CTTTACCGGCTCTGTGTCTCCGAGGTCTTTTGACAGACACCCGGGTGAGAGCCCTTC 138
DB 321 CGGTGGCAGATCTGTGTACGAGTTTCATGTGACAGACACCGCCCGCGGGGTGTG 380
QY 139 CCTATTACCGCGCTTTTGTGAGCGCTTTCCACCCCTTAAGACCCCTGGCCCGGGTTTC 198
DB 381 GGAATCTGGCGGACTGGGTGCGCGGTGCCACCGCTGGCCACCCGCAACGCGCAGC 440
QY 199 CTGGAAGAGGTCTTATGAGGTCTTGCAAGGAGCGGCTACTACCGCGCGGACACCTTC 258
DB 441 ACCGCCGATGTGTATAGCCCTTGCGGCAAGCTGGGTATCCAGAGGACCAAGCCCTTA 500
QY 259 CACCGCCCTGGCCCGGAAGCGTG-----GAGAGCTTCCCGGAGCTTTCGCC 303
DB 501 CACGAGTGGCGCACCGCTCATCGCCCGCAGACCAATGACGTGTGCGCAGATATCGAG 560
QY 304 GAGCTTCGGGGGCTTCTGTGTCTCGGCGCTTACACCGCGCGCGGTGGCTTCATCGCC 363
DB 561 ATCTGTGTACCTCTCGCGGCGCTCGGAGCTACACCGCGCGGTGGCGGTTCGTTCGT 620
QY 364 TTCCGGGAGCGGAGTGGCGGCTGGAACGGGAAAGTCCGAGGAGTCTTCCCGCTCTT- 422
DB 621 TACCGCCAGCGGGGTGCGGTGTGACACCAATGTGCGCGGTGTGGCTCGCGCGCTT 680
QY 423 -----CGCCCGGAAAAAGCCCAAGAGAAAGAGACTTTTCGCTTCGCCCAAGGC 471
DB 681 CACGCGCGCGCGACGCGCGGTGCGCATGCGTGCAGCGCAGACCAACGCTTTCGCG 740
QY 472 CTCCTCCCGGAGGGGTGACCCCGGGGGGTGTGAAACCAAGCCCTCATGAGACTCGGGGCC 531
DB 741 CTGTTCGCGCACCGGAGAGCGCGCTGAAATTTTCGTGCGCTGATGAGATTGGGTGG 800
QY 532 ACGGTCTGCGCTGCGCAAAACGGCCCGTTGCGGGGCGCTGCGCCCTTAGGGGCTTCTGCGCG 591
DB 801 ACGGTGTGACCGCGCCCGCACACCCCGGTGCGGGTTATGCGCGCTGATCTGTGCGCATGG 860
QY 592 GGAAGAGAGCGCCCGCGGGGCTTACCCCGCGCCAGAGACCGCGGCGAAGAGAGCGC 651
DB 861 CGGCAATGCGGGTTATCCGCGCTCGACGGTCC---GCCGCGCGGGGGGAGGCTTACACC 917
QY 652 CTCGTGCGCTCTGTCTCTCTCGGGCGG 678
DB 918 GGAACCGACCGCAAGTCCGCGGACCG 944

```

RESULT 12  
 US-10-437-963-35788  
 ; Sequence 35788, Application US/10437963  
 ; Publication No. US2004012343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Bouharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 35788  
 ; LENGTH: 1425  
 ; TYPE: DNA  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_39677C.1  
 US-10-437-963-35788

Query Match 8.8%; Score 86.2; DB 17; Length 1425;  
 Best Local Similarity 55.7%; Pred. No. 3.6e-11;  
 Matches 191; Conservative 0; Mismatches 143; Indels 9; Gaps 1;

```

QY 65 GGGGAGAGAGAGACCTTACCGCGCTTCTGTCTCCGAGGTCTTTCGACAGACCGG 124
DB 266 GCGAGAGAGAGAGGCGTACCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 325
QY 125 TGAAGAGCGCCCTCCCTATTACCGCGCTTTCGAGGCGTTTCCACCTGAAGGCG 184
DB 326 TACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 385
QY 185 TGGCGCGGCTTCTCTGAAAGGTCTTATGAGGTCTGCAAGGCGCGGCTACTACCGC 244
DB 386 TCGCGCGCGCACGCAAGAGAGGTGAACGATGTGGCGGGGCTCGGGTATTACCGGA 445
QY 245 GGGCGGAACACTCCACCGCGCTGCGCCGAAGGT-----GAGAGAGCTTCCCGCA 295
DB 446 GGGCTGTGATTTCTCTCGAGGAGCAAAACAAATTGTGAAGAGCGAGTTCCTTGC 505
QY 296 GCTTCCCGAGCTTGGGGGCTTCTGTCTCGGCGCTTACACCGCGCGCGGTGAGCT 355
DB 506 CAGCATCAACACTCGGTGAAGTTCTGTGATTTGGGGATTACAGCAGAGGCGATTGCTT 565
QY 356 CCATGCGCTTGGGAGCGGCTGCGCGCGGTGAGCGGGAACGT 398
DB 566 CCATAGCTTCAATGAGTTGTCTCTGTGTGTGTGAGCGGAATGT 608

```

RESULT 13  
 US-08-961-527-12/c  
 ; Sequence 12, Application US/08961527  
 ; Publication No. US2002003233A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Charles Kunich  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 391  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/961,527  
 ; FILING DATE:  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brookes, A. Anders  
 ; REGISTRATION NUMBER: 36,373  
 ; REFERENCE/DOCKET NUMBER: PB340P1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 12:  
 ; SEQUENCE CHARACTERISTICS:



Tue Jan 25 11:49:52 2005

us-09-938-901a-1.tmpb

LENGTH: 9909 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-12

Query Match 8.4%; Score 82; DB 6; Length 9909;  
Best Local Similarity 52.2%; Pred. No. 2.6e-10;  
Matches 218; Conservative 0; Mismatches 185; Indels 15; Gaps 1;

22 CTCCTGCGCTGTGTCACGGGAAAACGCCGCCCTCCCTCGCGGGGGAGAGAACCT 81  
5808 CTTCTGCGCTGTGTCACGGGAAAACGAGATTTCCTTGAGGAGAACTGAAAATCT 5749  
82 TACCGCGCTGTGTCACGGGCTCTCTGACGAGACCGGGGTGAGAGAGCCCTCCC 141  
5748 TATCATCTGCGGTATCTGAAATCATGCTTCAGCAGACGAGGTGATACAGTTATCCCT 5689  
142 TATTACCGCGCTGTGTCACGGGCTCTTCCACCCCTGAAAGCCCTGCGCGGCTTCCTG 201  
5688 TACTACGAAAGATTTTGTGAGCTGTTTCACTGTCGAAAGTCTGCACTGCGCTGAG 5629  
202 GAAAGAGTCTTAGGCTCTGCGAGGCGGGGCTA-----CTACCGCGG 246  
5628 GAGAGTTACTGAAACCTTGAGGAGCTTGAGCTTATTTCTGAGTTCCGAAATATGAG 5569  
247 GCGGAACACCTCCACCGCTGCGCGAGAGCTGAGAGAGCTTCCCGAGCTTGCGCGAG 306  
5668 GCTGACGCCAGAGATTTATGACTGACTTGTGTGCAATTTCCAAATCTATGAAAGGA 5509  
307 CTTGCGGCGCTTCTGCTGCTCGGCGCTTACACCGCGCGCGGCTTCATGCGCTTC 366  
5508 ATTTCAGCTTGAAAGGAGATTGACCTTACAGCAGAGAGCATTTCCAGATTTGCTTTT 5449  
367 GGGAGCGGCTGCGCGGCTGACGAGAGCTCCGAGAGGCTCTCCCGCTCTTCG 424  
5448 AACTTGCTGAGCAGCTGTAGATGTATGTATGATGCGGCTCTTGCGCTGTTTG 5391

RESULT 14  
US-10-158-844-12/c  
Sequence 12, Application US/10158844  
Publication No. US20040029118A1  
GENERAL INFORMATION:  
APPLICANT: Kunsch et al.  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-R  
COMPUTER: Dell Latitude Pentium 3  
OPERATING SYSTEM: Windows 98  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/158,844  
FILING DATE: 03-Jun-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/961,527  
FILING DATE: 1997-10-30  
APPLICATION NUMBER: US 60/029,960  
FILING DATE: 1996-10-31  
ATTORNEY/AGENT INFORMATION:  
NAME: Hyman, Mark J.  
REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB340P1D1  
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:  
LENGTH: 9909 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 12;  
US-10-158-844-12

Query Match 8.4%; Score 82; DB 16; Length 9909;  
Best Local Similarity 52.2%; Pred. No. 2.6e-10;  
Matches 218; Conservative 0; Mismatches 185; Indels 15; Gaps 1;

22 CTCCTGCGCTGTGTCACGGGAAAACGCCGCCCTCCCTCGCGGGGGAGAGAACCT 81  
5808 CTTCTGCGCTGTGTCACGGGAAAACGAGATTTCCTTGAGGAGAACTGAAAATCT 5749  
82 TACCGCGCTGTGTCACGGGCTCTCTGACGAGACCGGGGTGAGAGAGCCCTCCC 141  
5748 TATCATCTGCGGTATCTGAAATCATGCTTCAGCAGACGAGGTGATACAGTTATCCCT 5689  
142 TATTACCGCGCTGTGTCACGGGCTCTTCCACCCCTGAAAGCCCTGCGCGGCTTCCTG 201  
5688 TACTACGAAAGATTTTGTGAGCTGTTTCACTGTCGAAAGTCTGCACTGCGCTGAG 5629  
202 GAAAGAGTCTTAGGCTCTGCGAGGCGGGGCTA-----CTACCGCGG 246  
5628 GAGAGTTACTGAAACCTTGAGGAGCTTGAGCTTATTTCTGAGTTCCGAAATATGAG 5569  
247 GCGGAACACCTCCACCGCTGCGCGAGAGCTGAGAGAGCTTCCCGAGCTTGCGCGAG 306  
5668 GCTGACGCCAGAGATTTATGACTGACTTGTGTGCAATTTCCAAATCTATGAAAGGA 5509  
307 CTTGCGGCGCTTCTGCTGCTCGGCGCTTACACCGCGCGCGGCTTCATGCGCTTC 366  
5508 ATTTCAGCTTGAAAGGAGATTGACCTTACAGCAGAGAGCATTTCCAGATTTGCTTTT 5449  
367 GGGAGCGGCTGCGCGGCTGACGAGAGCTCCGAGAGGCTCTCCCGCTCTTCG 424  
5448 AACTTGCTGAGCAGCTGTAGATGTATGTATGATGCGGCTCTTGCGCTGTTTG 5391

RESULT 15  
US-09-864-866-46  
Sequence 46, Application US/09864866  
Patent No. US20020127656A1  
GENERAL INFORMATION:  
APPLICANT: Lloyd R. Stephen  
APPLICANT: McCullough, Amanda K.  
TITLE OF INVENTION: DNA REPAIR POLYPEPTIDES AND METHODS OF USE  
FILE REFERENCE: 265, 001/0101  
CURRENT APPLICATION NUMBER: US/09/864,866  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/206,279  
PRIOR FILING DATE: 2000-05-23  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 46  
LENGTH: 968  
TYPE: DNA  
ORGANISM: Micrococcus luteus  
US-09-864-866-46  
Query Match 8.3%; Score 81.2; DB 9; Length 968;  
Best Local Similarity 47.2%; Pred. No. 5.8e-10;  
Matches 300; Conservative 0; Mismatches 318; Indels 18; Gaps 1;  
44 ACGCCCGCCCTCCCTGCGGGGGAGAGACCTTACCGCGCTGCTGCTGAG 103  
200 ACCGTACCGCTGCGCGAGCTGACTTCGAGAGCCGCTTCACTGCTGAGCG 259  
104 TCTTCTGAGGAGACCGGCTGAGAGCGCTTCCTTATACCGCGCTTCTGAGC 163



Db 260 TGCTGTCCGCCAGACACCGAGCTGCGGGTCAACGACGACCGCGGCGCTGTCGCC 319  
QY 164 GCTTTCCACCCCTGAAGGCCCTTGCCCGCGGCTTCCCTGGAAGAGTCTTAGGCTGCGC 223  
Db 320 GCTTCCCGATGCCCCGATGACCGCGGCAACGAGCCGACGCTGACAGAGCTGTC 379  
QY 224 AGGGGCGGGCTACTACCGG-----CGGGCGAACAACCTCACCGCC 265  
Db 380 GCTCCACGGGGTTCTACCGGAACAAGGCTCCCGGATCTGCGGCTGTCCAGAGTCTG 439  
QY 266 TGGCCCGAAGCGTGAAGAGCTTCCCGGAGCTTCCCGGAGCTTCCGGGGCTTCTGTC 325  
Db 440 TGGCCCGGACGACGCGGAGAGTCCCGCGCGTCTCGAGGACTGTGCGGCTGCGCGGG 499  
QY 326 TCGGGCTTACACCGGGCGCGCGTGCCTTCATCGCTTCCGGGAGCGGAGTGGCGCG 385  
Db 500 TGGCCCGCAAGACCGGCTTCTGTGTGCTCGGCAACGCTTCGGCCAGCCCGGATCACCG 559  
QY 386 TGGACGGGAACGTCCGGAAGGCTCTCTCCGCTCTTCCGCGGGAAGCCCAAGGAGA 445  
Db 560 TGGACACGCACTTCGCGCGGCTGCGCGGCGCTTGGGGTTACGAGAGAGACCGCGG 619  
QY 446 AAGAGCTTTGCGCTTCGCCAGAGGCTCTCCCGAGAGGCGTGAACCCGGGGTGTGA 505  
Db 620 GTAAAGTGCAGACGCGCGTGGCGCGCTGTTCCTCCCGCGCGGACTGGACGATGCT 679  
QY 506 ACCAGGCTCTCATGAGCTTCGGGGCCACGCTTGCCTGCCGAAAAGCGCCCGTTGCGGG 565  
Db 680 CCCACCGGCTGATCTTCCACGCGCGCGCTGTGCCACGCGCGCGCGCTGCGGCGC 739  
QY 566 CTTGCCCTTAAGGGGCTTCTGCGGGGGAAGAGGCCCGGGCGCTACCCGCGGCCA 625  
Db 740 GGTGCCGATCGCGCGTGTGCTCCGTCTACCGCGGGGGAGACCGACCCGAGCGGG 799  
QY 626 GGAAGCGCGCGGGAAGAGAGAGCGCTGTCGCC 661  
Db 800 CGCGCGCTGCTGCGCTACGAGCTCAAGCCGCGC 835

Search completed: January 20, 2005, 06:10:40  
Job time : 678 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 20, 2005, 04:37:51 ; Search time 26 Seconds  
(without alignments)  
828.975 Million cell updates/sec

Title: US-09-938-901a-2  
Perfect score: 1701  
Sequence: 1 MEAMRKALLAWYRENNARPLP.....VLKRLPLLAHAGVPLPDA 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

## Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	505	29.7	403	4 US-09-107-532A-7308	Sequence 7308, Ap
2	485	28.5	395	4 US-09-134-000C-5115	Sequence 5115, Ap
3	481.5	28.3	535	3 US-08-813-574-2	Sequence 2, Appl
4	477.5	28.1	470	4 US-09-252-991A-23310	Sequence 23310, A
5	477	28.0	375	4 US-09-198-452A-421	Sequence 421, App
6	471.5	27.7	350	3 US-08-813-574-9	Sequence 9, Appl
7	471.5	27.7	350	3 US-09-651-656-37	Sequence 37, Appl
8	471.5	27.7	350	3 US-09-650-855-37	Sequence 37, Appl
9	458.5	27.0	391	4 US-09-583-110-4583	Sequence 4583, Ap
10	457.5	26.9	516	4 US-09-489-039A-11959	Sequence 11959, A
11	450	26.5	347	4 US-09-543-681A-5315	Sequence 5315, Ap
12	443	26.0	360	3 US-09-134-001C-4684	Sequence 4684, Ap
13	442	26.0	346	4 US-09-328-352-4982	Sequence 4982, Ap
14	389.5	22.9	188	4 US-09-629-222A-36	Sequence 36, Appl
15	355	20.9	414	4 US-09-540-236-2578	Sequence 2578, Ap
16	274.5	16.1	221	1 US-08-663-023-17	Sequence 17, Appl
17	274.5	16.1	221	4 US-09-402-959A-10	Sequence 10, Appl
18	274.5	16.1	225	4 US-10-037-927B-14	Sequence 14, Appl
19	240.5	14.1	188	4 US-09-629-222A-34	Sequence 34, Appl
20	199	11.7	188	4 US-09-864-866-43	Sequence 43, Appl
21	191	11.2	185	4 US-09-629-222A-35	Sequence 35, Appl
22	188	11.1	227	4 US-09-107-532A-3843	Sequence 3843, Ap
23	176	10.3	207	2 US-08-808-550-36	Sequence 36, Appl
24	176	10.3	259	2 US-08-808-550-39	Sequence 39, Appl
25	172	10.1	209	2 US-08-808-550-31	Sequence 31, Appl
26	171	10.1	204	2 US-08-808-550-32	Sequence 32, Appl
27	164.5	9.7	211	2 US-08-808-550-37	Sequence 37, Appl

28	164.5	9.7	297	2 US-08-808-550-42	Sequence 42, Appl
29	164.5	9.7	304	2 US-08-808-550-2	Sequence 2, Appl
30	164	9.6	210	4 US-09-710-279-972	Sequence 972, Appl
31	164	9.6	224	4 US-09-710-279-2684	Sequence 2684, Ap
32	160	9.4	224	3 US-09-134-001C-3225	Sequence 3225, Ap
33	159	9.3	209	4 US-09-583-110-4518	Sequence 4518, Ap
34	158	9.3	185	4 US-09-134-000C-3440	Sequence 3440, Ap
35	154	9.1	213	2 US-08-808-550-35	Sequence 35, Appl
36	145	8.5	207	2 US-08-808-550-29	Sequence 29, Appl
37	145	8.5	211	2 US-08-808-550-38	Sequence 38, Appl
38	145	8.5	211	3 US-09-651-656-25	Sequence 25, Appl
39	145	8.5	211	3 US-09-650-855-25	Sequence 25, Appl
40	145	8.5	211	4 US-09-492-709A-296	Sequence 296, Ap
41	142	8.3	216	4 US-09-489-039A-9560	Sequence 9560, Ap
42	137.5	8.1	184	4 US-09-629-222A-33	Sequence 33, Appl
43	134	7.9	207	2 US-08-808-550-30	Sequence 30, Appl
44	128	7.5	628	4 US-09-252-991A-30904	Sequence 30904, A
45	127.5	7.5	212	4 US-09-252-991A-17572	Sequence 17572, A

## ALIGNMENTS

RESULT 1  
US-09-107-532A-7308  
Sequence 7308, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107, 532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085, 598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneko  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-8277  
TELEFAX: (781)893-5007  
INFORMATION FOR SEQ ID NO: 7308:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 403 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...403  
SEQUENCE DESCRIPTION: SEQ ID NO: 7308:  
US-09-107-532A-7308

Query Match 29.7%; Score 505; DB 4; Length 403;  
Best Local Similarity 32.7%; Pred. No. 2.3e-42;  
Matches 118; Conservative 66; Mismatches 113; Indels 64; Gaps 11;  
QY 9 LAWRENAPRLPMRGKEDPYRVLSEVLLQOTRVEQALPYRRPRLERFPTLKALAAASLE 68  
DB 32 ICHYEOKRNLPMRYNDPFRIMWSEIMLQOTRVDYIDFYRMEWEPFLIBELANPEE 91  
QY 69 EVLRVWOGAGYRRRAEHLRLARSV-----BELPPSAELRGDLGCPYPAVAASIAFG 123  
DB 92 KLIKAMEGLGYSSRAANIQAAKQIMSEFDGEMFQTEIEISLKGICPYTTGAIASIAFG 151  
QY 124 ERVAUVCNARVLSRLFARES-----PKKELFALAGLLPEGVPGVMQALMELGATV 179  
DB 152 LPEPAVDGNVNRVLSRLFCIEADIAKASSRKIFDEARKIIDEHNPGEFQAMMDLGSAT 211  
QY 180 CLPKRPGCAGPLGAF-----RGKEAPGRYPAPRKRRAKEERLVALVLGRKGVH----- 230  
DB 212 CTPTSPKCTCPTIOAFCLANKRGQT--SFPVKTKKAKPDVYIISALQNHSGAYFEE 269  
QY 231 -----LERLEGRFOGLYGVPLF-----PPEELP-GREAAFGV 261  
DB 270 RDSQKLLANMTFPMMEVTOEYERLKKEMETKEIDLFDDLVAEEQNLPFEKQELFW 329  
QY 262 RSRRLGEVRHALTRRLRLEV-RGALMEGEGD-----PKKR-----LPKLMKV 306  
DB 330 QTRHLGEVTHIFSHLKHVLLFYGRATEGAQEFTENKSKMLPDAFDSVFPKVMKL 389  
QY 307 L 307  
DB 390 V 390

RESULT 2  
US-09-134-000C-5115  
; Sequence 5115, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134.000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5115  
; LENGTH: 395  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-5115

Query Match 28.5%; Score 485; DB 4; Length 395;  
Best Local Similarity 33.5%; Pred. No. 2.3e-40;  
Matches 115; Conservative 56; Mismatches 108; Indels 64; Gaps 8;  
QY 2 EAMRKA-----LLAWYRENARPLPMRGKEDPYRVLSEVLLQOTRVEQALPYRRFL 53  
DB 9 EAWSPAKYSSFOEDFLAWYERKKNLPMRANTDAYRIMISIMLQOTRVDYIDFYFRFM 68  
QY 54 ERFPFLKALAAASLEEVLRVWOGAGYRRRAEHLRLARSV-----BELPPSAELRGDLG 108  
DB 69 EMFPTIQLAEAPDCKLIKAMEGAGYSSRAANLKYAAQIVSEFGKMPDITIEDIRSLKG 128  
QY 109 LGPRTAAVAASIAGERVAADVGNVRLSLFARES-----PKKELFALAGLLPEGVD 164  
DB 129 IGPYTAAGISIAINLEPAIDGNVNRVLSRLFCIEDIAKASSRKIFEAAMLKIIDRER 188  
QY 165 PGVMQALMELGATVCLPKRPGCAGPLGAFCRGKEAP--GRYPAPRRK-RAKEERLVAL 221

DB 189 PGDFNQALMDLGSAVCTPTSPKCESCPLOQYCAAYOADRMATVAVPKSKKVPKEDVYVGT 248  
QY 222 VLLGRGVNLELEGRFQG--LYGVPLPPEELPGE----- 256  
DB 249 ILENKQOEFL--LEQRPETGLANMMLPFEIESISKQFOQLKLAOPATEKQTLIELBP 306  
QY 257 -----AAFGVSRPLGEVRHALTRRLRLEV 282  
DB 307 VNEPLVAEPPVNFMDYETVWVQKRTLGEVNHIFSHLKHILV 349

RESULT 3  
US-08-813-574-2  
; Sequence 2, Application US/08813574  
; Patent No. 6033473  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Ying-Fei  
; TITLE OF INVENTION: Human Muty  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: IBM Compatible  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/813,574  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/013,132  
; FILING DATE: 11-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Han, William T  
; REGISTRATION NUMBER: 34,344  
; REFERENCE/DOCKET NUMBER: ATG50002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-5219  
; TELEFAX: 610-270-4026  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 535 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-813-574-2

Query Match 28.3%; Score 481.5; DB 3; Length 535;  
Best Local Similarity 36.5%; Pred. No. 8e-40;  
Matches 135; Conservative 46; Mismatches 106; Indels 83; Gaps 13;  
QY 3 AMRKALLAWYRENARPLPMRG-----EKDPYRVLSEVLLQOTRVEQALPYRRFLE 54  
DB 81 AFGSLLSWVDQKRLPMRRRAEDMDLDRAYAVWVSEVMLQOTVAVINYTGMMQ 140  
QY 55 RFPFLKALAAASLEEVLRVWOGAGYRRRAEHLRLARS-VEEL---PPPSAEL--RGDLG 108  
DB 141 KMPTLDDLASASLEEVNQJLWAGLVYSSRRRLQEGARKVBEIGMPTAETLQQLPG 200  
QY 109 LGPRTAAVAASIAFGERVAAVDGNVNRVSLFA-----RESPKELFALAGLLPEGV 163  
DB 201 VGRYTGALASISIFGQATGVVDGNVNRVLCRVRAIADSSSTIVSQGLAGLQQLV-DPA 259  
QY 164 DPGVMQALMELGATVCLPKRPGCAGPLGAFCRGKE-----A 201

Db 260 RPDGPNQAMELGATVCTPQRPLCSQCPEVSLCRARQVREQOLLASGSLSSPDVEEKA 319  
Qy 202 P-----GRYPARKRAK---BERLVALV-----LGRGVHLER- 233  
Db 320 PUTGQCHLCPPEEPDQTLGVNPPRKASRKPRRESSATCYLEOPGALQILLVQRP 379  
Qy 234 LRGFQGLYVPLPPPEELPGRBAF-----GVRSRPLGEVNHATHRLRY 280  
Db 380 NSGLLAGLWBFPEVTEPESQOLRKALLQELQWAGPLPATHILHGEVHTSHIKTY 439  
Qy 281 EVRGALWEEB 290  
Db 440 QVYGLALEGQ 449

RESULT 4  
US-09-252-991A-23310  
; Sequence 23310, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23310  
; LENGTH: 470  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23310

Query Match 28.1%; Score 477.5; DB 4; Length 470;  
Best Local Similarity 36.2%; Pred. No. 1,7e-39;  
Matches 126; Conservative 58; Mismatches 119; Indels 45; Gaps 14;

Qy 2 EAMRKALLAMY-RENARPLPWRGKDPYRVLVSEVLLQOTRVEQALPYRRFLERPTLK 60  
Db 119 EGFNGAVLMDWRHGRKDLPMOQGIPIYVWVSEIMLQOTQVSTVGYDRFMALPDVE 178  
Qy 61 ALAASLEEVLRWQAGYRRRAEHLRLA-----RSVEELPPSPALRGPLGPGYTA 115  
Db 179 ALAAAEDVHLMTGLGYSSRARNHKTQAQIVERRHAGEPRPDVQLAELPGIGSTAG 238  
Qy 116 AVASIAFGERRVAANGVNRVLSRLFARE-----SPK-EKELPALAGLPEGVDPGVNQ 170  
Db 239 AIASISMGIRAPLDGNVRVLAARYLAQOGYGPSPVAPALWMAARFTPH-ARVNHVYQ 297  
Qy 171 ALMELGATVCLPAPRCGACPLGAFCR-----GKEAPGRYPAPRKRA-KEERLVALVLG 225  
Db 298 AMMDLGAITCTSKPSCCLCPVSGRAHLGREAL-YQPPRKALQOKRTLMITLAN 355  
Qy 226 RKG---VHLERLEGRFQGLYVPLPPPEELPGRBAFGVRSRLGEVR-----HALTTR 276  
Db 356 RQGAILLYRRPSSGLWGLMSLPRL--DDLGLPLAARSHSLALGERRELSGLTHTFSHP 413  
Qy 277 RLRVE-----VRGA-----LWGEGBEDPWKRRLPLKMEVYLAKA 310  
Db 414 QLAIEPWLVAVEGAPRAVAEGDWLWNLATPP-RIGLAAPVKGLAKRA 460

RESULT 5  
US-09-198-452A-421  
; Sequence 421, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Griffiths, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
; TITLE OF INVENTION: and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 421  
; LENGTH: 375  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-421

Query Match 28.0%; Score 477; DB 4; Length 375;  
Best Local Similarity 38.6%; Pred. No. 1.4e-39;  
Matches 117; Conservative 44; Mismatches 112; Indels 30; Gaps 7;

Qy 6 KALLAWYRENARPLPWRGKDPYRVLVSEVLLQOTRVEQALPYRRFLERPTLKALAAA 65  
Db 22 EALKKVFEEKKSLPFRDNPPTYSVWVSEVMLQOTRAEVIDYFNQMMERFPTESLAA 81  
Qy 66 SLEEVLRWQAGYRRRAEHLRLASVE-----LPPSPALRGPLGPGYTAANVASI 120  
Db 82 KEEDVIKMEGIGYSSRAHLLEGARWMBEPFGKIIPDAISLAQIRGVGPYVHAIIAF 141  
Qy 121 AFGERVAADVGNVNRVLSRLFARESPEKE-----LFALAQGLPEGVDPGVNQALMEL 175  
Db 142 AKRRRAADVGNVNRVLSRLFILETISIDLESTRTWRSRAQALLPR-KSPREVAEHLIEL 200  
Qy 176 GATVCLPKPRPGACPLGAFCRGKEAPGRYPAPRKRAKE-----ERLVALVL-GRGVY 230  
Db 201 GACIC-KKVPOCHRCVQACGAMWENKQFVLPVRARKKVIFLRLVALVLYDGLVYE 259  
Qy 231 LERLEGRFQGLYVPLPPPEELPGRBAFGVRSR-----PLGEVNHATHRR 277  
Db 260 KRPRKEMAGLYEFPYIEVEPEEGLODIEGFTKQWELSLSPLEFLGNLKEORHAPTNRK 319  
Qy 278 LRV 280  
Db 320 VHL 322

RESULT 6  
US-08-813-574-9  
; Sequence 9, Application US/08813574  
; Patent No. 6013473  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Ying-Fei  
; TITLE OF INVENTION: Human Muty  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/813,574  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/013,132  
; FILING DATE: 11-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Han, William T  
; REGISTRATION NUMBER: 34,344  
; REFERENCE/DOCKET NUMBER: ATG50002  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5219  
TELEFAX: 610-270-4026  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 350 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-813-574-9

Query Match 27.7%; Score 471.5; DB 3; Length 350;  
Best Local Similarity 35.1%; Pred. No. 4.5e-39;  
Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAWYRENAR-PLPWGKEDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66  
DB 10 VLDWYDKYGRKTLPMQIDKTYKMWLSEVMLQOTQVATVIFYFERFMARFPTVLDLANAP 69  
QY 67 LEEVLRVMOAGGYRRAHRLHRLARSVEEL-----PPSFAELRGLPGIGPYTAANVASIA 121  
DB 70 LDEVHLMTGTGAGYARARNLHKAQOAVATLHGKFPETFEVVALPGVGRSTAGAILSLIS 129  
QY 122 FGERVAADVGNVRRVLSRLFARES-PKEKE-----LFALAQGLPE-GVDPGVNQAAMLEL 175  
DB 130 LQKHFPILDGNVRRVLCVAVSGWPGKKEVENKMSLSEQVTAVER--FNQAAMDL 187  
QY 176 GATVCLPRPRGACPL--GAFCKGKEAPGRYPAPRRKRAKEERLVALVLLGRGVHL-- 231  
DB 188 GAMICTRSKPKCSLCPLONGCIAANNSWALYPGKKPKQTLPERTGYFLLQHEDEVLLA 247  
QY 232 -ERLEGRFOGLYGVPLFPPEE-----LPGEAAGVRSRPLGEVRHALTHRLRV----- 280  
DB 248 QRPSPGLMGGLYCPQPADEESLRQWLAQROIADNLTO-LTAFRHTFSHFHLDIVPMWL 306  
QY 281 -----EVRGALWEGEGEDP--WKRPPLKMEKVLKALPL 313  
DB 307 PVSSFTGCMDEGNALWYNLAQPPSVGLAAPVERLLQO-LRTGAPV 350

## RESULT 7

US-09-651-656-37  
Sequence 37, Application US/09651656  
Patent No. 6340566  
GENERAL INFORMATION:  
APPLICANT: MCCUTHEN-MALONEY, SANDRA  
APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY  
TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE  
TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,  
TITLE OF INVENTION: DNA DAMAGE AND DNA MISMATCHES  
FILE REFERENCE: IL-10689  
CURRENT APPLICATION NUMBER: US/09/651,656  
CURRENT FILING DATE: 2000-08-29  
PRIOR APPLICATION NUMBER: 60/192,764  
PRIOR FILING DATE: 2000-03-28  
NUMBER OF SEQ ID NOS: 106  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 37  
LENGTH: 350  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-651-656-37

Query Match 27.7%; Score 471.5; DB 3; Length 350;  
Best Local Similarity 35.1%; Pred. No. 4.5e-39;  
Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAWYRENAR-PLPWGKEDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66  
DB 10 VLDWYDKYGRKTLPMQIDKTYKMWLSEVMLQOTQVATVIFYFERFMARFPTVLDLANAP 69  
QY 67 LEEVLRVMOAGGYRRAHRLHRLARSVEEL-----PPSFAELRGLPGIGPYTAANVASIA 121

DB 70 LDEVHLMTGTGAGYARARNLHKAQOAVATLHGKFPETFEVVALPGVGRSTAGAILSLIS 129  
QY 122 FGERVAADVGNVRRVLSRLFARES-PKEKE-----LFALAQGLPE-GVDPGVNQAAMLEL 175  
DB 130 LQKHFPILDGNVRRVLCVAVSGWPGKKEVENKMSLSEQVTAVER--FNQAAMDL 187  
QY 176 GATVCLPRPRGACPL--GAFCKGKEAPGRYPAPRRKRAKEERLVALVLLGRGVHL-- 231  
DB 188 GAMICTRSKPKCSLCPLONGCIAANNSWALYPGKKPKQTLPERTGYFLLQHEDEVLLA 247  
QY 232 -ERLEGRFOGLYGVPLFPPEE-----LPGEAAGVRSRPLGEVRHALTHRLRV----- 280  
DB 248 QRPSPGLMGGLYCPQPADEESLRQWLAQROIADNLTO-LTAFRHTFSHFHLDIVPMWL 306  
QY 281 -----EVRGALWEGEGEDP--WKRPPLKMEKVLKALPL 313  
DB 307 PVSSFTGCMDEGNALWYNLAQPPSVGLAAPVERLLQO-LRTGAPV 350

## RESULT 8

US-09-650-855-37  
Sequence 37, Application US/09650855  
Patent No. 6365355  
GENERAL INFORMATION:  
APPLICANT: MCCUTHEN-MALONEY, SANDRA  
APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY  
TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA  
TITLE OF INVENTION: MISMATCHES  
TITLE OF INVENTION: DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA  
FILE REFERENCE: IL-10284  
CURRENT APPLICATION NUMBER: US/09/650,855  
CURRENT FILING DATE: 2000-08-29  
PRIOR APPLICATION NUMBER: 60/192,764  
PRIOR FILING DATE: 2000-03-28  
NUMBER OF SEQ ID NOS: 106  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 37  
LENGTH: 350  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-650-855-37

Query Match 27.7%; Score 471.5; DB 3; Length 350;  
Best Local Similarity 35.1%; Pred. No. 4.5e-39;  
Matches 121; Conservative 61; Mismatches 120; Indels 43; Gaps 13;

QY 8 LLAWYRENAR-PLPWGKEDPYRVLVSEVLLQOTRVEQALPYRRFLERFPTLKALAAS 66  
DB 10 VLDWYDKYGRKTLPMQIDKTYKMWLSEVMLQOTQVATVIFYFERFMARFPTVLDLANAP 69  
QY 67 LEEVLRVMOAGGYRRAHRLHRLARSVEEL-----PPSFAELRGLPGIGPYTAANVASIA 121  
DB 70 LDEVHLMTGTGAGYARARNLHKAQOAVATLHGKFPETFEVVALPGVGRSTAGAILSLIS 129  
QY 122 FGERVAADVGNVRRVLSRLFARES-PKEKE-----LFALAQGLPE-GVDPGVNQAAMLEL 175  
DB 130 LQKHFPILDGNVRRVLCVAVSGWPGKKEVENKMSLSEQVTAVER--FNQAAMDL 187  
QY 176 GATVCLPRPRGACPL--GAFCKGKEAPGRYPAPRRKRAKEERLVALVLLGRGVHL-- 231  
DB 188 GAMICTRSKPKCSLCPLONGCIAANNSWALYPGKKPKQTLPERTGYFLLQHEDEVLLA 247  
QY 232 -ERLEGRFOGLYGVPLFPPEE-----LPGEAAGVRSRPLGEVRHALTHRLRV----- 280  
DB 248 QRPSPGLMGGLYCPQPADEESLRQWLAQROIADNLTO-LTAFRHTFSHFHLDIVPMWL 306  
QY 281 -----EVRGALWEGEGEDP--WKRPPLKMEKVLKALPL 313  
DB 307 PVSSFTGCMDEGNALWYNLAQPPSVGLAAPVERLLQO-LRTGAPV 350

## RESULT 9

```
US-09-583-110-4583
; Sequence 4583, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4583
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-4583

Query Match
Best Local Similarity 33.8%; Score 458.5; DB 4; Length 391;
Matches 122; Conservative 53; Mismatches 127; Indels 59; Gaps 11;

QY 3 AMRKALLAWYRENARPLPMRGEKDPYRVLVSEVLLQOTRYEQALPYRRFLERFPTLKAL 62
DB 19 SFEKRLAWYDENKRDLPWRRSKNPIHWSEIMLQOTRYDVIPIYERFLDWEPTVESL 78
QY 63 AAASLEEVLRWOGAGYRRAEHLRLARSV-----ELPPSFAELRGDLPGPYTAAAY 117
DB 79 ACAPEESLLKAEGLGYRVRWMOAAQOIMTDFGQFPNTYEGISLKGIGPYTAGAI 138
QY 118 ASIAPGERVAADVGNVRVLSRLFARES-----PKERELPALAQGLPEGVDPGVNQAALM 173
DB 139 SSIAPRLPPAYDGNVRLARLAFEVNHDIGIPSNRKIQAMMEILINPRDEDFQALM 198
QY 174 ELGATVCLPKRPGCAGPLGAFRCGE--APGRYP--APRRKRAKEERLVALVLLGRKGV 229
DB 199 DLGSDIESPVNPRPESPYVDFAAYQNGTMDRYPIKSPKKKV--PIYLKALVVKNSQGG 257
QY 230 HL---ERLGRFOGLYGVLP-----FPPBE--LPGRBAFGVRSPLGE----- 268
DB 258 FLEKNSEKRLAGFWHPFPIEVNDFSQBEQDPLFQVABESVNFSPSPESFOQDYDLD 317
QY 269 -----VHALTHRLRLRYEV-----RGALWEGEGEDPWRKRLPKLMEK 305
DB 318 VMDLVCFPTQVHVSFRKMHVQIVAGOVSDPHDFSDREVRLSP--EEFKNTPLAKPOOK 376
QY 306 V 306
DB 377 I 377

RESULT 10
US-09-489-039A-11959
; Sequence 11959, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 11959
; LENGTH: 516
; TYPE: PRT
```

```
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11959

Query Match
Best Local Similarity 34.8%; Score 457.5; DB 4; Length 516;
Matches 120; Conservative 63; Mismatches 119; Indels 43; Gaps 14;

QY 8 LLAWEENAR-PLPMRGEKDPYRVLVSEVLLQOTRYEQALPYRRFLERFPTLKALAAAS 66
DB 176 VLDWYDKGRKTLPMQIATPYKWLSEVLLQOTVTVIPIYERFLDWEPTVESL 235
QY 67 LEEVLRWOGAGYRRAEHLRLARSV-----ELPPSFAELRGDLPGPYTAAAYASIA 121
DB 236 LDEVHLMTGLGYRRAEHLRLARSV-----ELPPSFAELRGDLPGPYTAAAYASIA 295
QY 122 FEEVVAADVGNVRVLSRLFA-----RESPKEELPALAQGLP--EGVDPGVNQAALMBL 175
DB 296 LGOHYPILDGNVRLARCYAASGMPGKKEVERKRLMDISEEVTPAEGER--FNQAMMDL 353
QY 176 GATVCLPKRPGCAGPL--GAFRCGEARGRYPARRKRAKEERLVALVLLGRKG--VHL 231
DB 354 GAMVCTRSPKCELCPLSNGCVAAVYANHSWAETPGKKPKOTLPER--TGYPFLMQHGDVFL 412
QY 232 ERLE--GRFOGLYGVLPFPPPE--ELPGRBAFGVRSPLGEV--RHATHRLRLV----- 280
DB 413 SQRPVGLMGGLFCFPQFADAEALREWLAKQRIKADNLVQLTAFRTTFSHFLDIYPMML 472
QY 281 -----EVGALWEGEGEDP--WKRPLPKLMEKVLRLKALPL 313
DB 473 TVHSSGAYMDEGNALMYNLAPSPSVGLAAPVERLLQO--LKAGAPV 516

RESULT 11
US-09-543-681A-5315
; Sequence 5315, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5315
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5315

Query Match
Best Local Similarity 33.2%; Score 450; DB 4; Length 347;
Matches 115; Conservative 64; Mismatches 123; Indels 44; Gaps 12;

QY 2 EAMRKALLAWYRENAR-PLPMRGEKDPYRVLVSEVLLQOTRYEQALPYRRFLERFPTLK 60
DB 6 QOFSQVLDWYHKKYGRKTLPMQOKEKTPYHVMVSEVLLQOTVATVIPIYERFLARPPNV 65
QY 61 ALAASLEEVLRWOGAGYRRAEHLRLARSV-----ELPPSFAELRGDLPGPYTAA 115
DB 66 ALAKAPLDEVHLMTGLGYRRAEHLRLARSV-----ELPPSFAELRGDLPGPYTAA 125
QY 116 AVASTAFGEVVAADVGNVRVLSRLFARES--PKEXE--LALAQGLP--EGVDPGVN 169
DB 126 AILSLSKRPYPILDGNVRLARCYAVGSGKKEVERKRLMDISEEVTPTGCV--YFN 183
QY 170 QALMBELGATVCLPKRPGCAGPLGAF-----RGKEARGRYPARRKRAKEERLVALV 223
DB 184 QAMMDLGAMVCTRSPKCELCPLNGLCIAYAHSMADYGGKP--KKAIPKTYFLIL 240
QY 224 LGRKGVHLERL--EGRFQGLYGVLPFP-----ELPGRBAFGVRSPLGEVRLHATHRR 277
```

Db 241 QYDNLVWLDRPPAGIWMGGIFAFPOPEFTYMLLEQWLAHGHJENSTBEDLLSFRATFSFPH 300

Qy 278 LRV-----EVRGALMEGEGEDPWKRPDKPLKIMEKYLRK 309

Db 301 LDIVPICVKSTFTSLMEAGQIGIWNYL-QTPASVGIAPVENVTLNQ 345

```

RESULT 12
US-09-134-001C-4684
; Sequence 4684, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4684
; LENGTH: 360
; TYPE: prt
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4684

```

Query Match	26.0 %	Score 443 ;	DB 3 ;	Length 360 ;
Best Local Similarity	34.3 %	Pred. No. 3,4e-36 ;		
Matches 103 ;	Conservative 56 ;	Mismatches 119 ;	Indels 22 ;	Gaps 7 ;
QY				
3	AMRKALLAVYRENARPLPMRGKEDPYRVLYSEVLLQOTVEQALPYRRREFLRPFPLTKL	62		
	::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::			
19	SFKQIEDIEFHKKOROMPMEKNTETNPYYIMLSEVMLOOTOONTVIDYYFIRHREFITQSL	78		
QY				
63	AAASLEEVLRVMOGAGYVRRRAEHLHLASVSE-----ELPPSPAEIRLGGPGIGPYTAAY	117		
	::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::			
Db	79	SEANDEBVLKYMGGGLGYSPARNPHTAVKENVNNNYDGEVYDDEPSFKKLKGVPYTOAAV	138	
QY				
118	ASIAFGERRVAAVUNGNNRVLSRL--PARSPKKEKELFALAQGLPEGV--DPGVWNOALME	174		
	::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::			
139	MSIAFHNHPLATVGNVFNWMSRLNNDYRDKLOSTRAPFQELHPVLYKADGIFNDQAME	198		
Db				
175	LGATVCLPCKPRCGACPLGAFCRGKEAPGYYPAFRKRAKEERLV--ALVLLGRGGVHL	231		
	::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::			
199	LGALVTCPRKPLCLFCPIQSHCEAFHMGTOQLPVTKTSJNKKTIEQKVELINDNGOYL	258		
QY				
232	--EELBGRFQGLVQPLRPPEELDGRRA-----AFGVSRPLGVERHALLTHRLRLEY	282		
	::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::			
259	LEKKEKELKLNMGWFFM--REGTANADVIGDDLOKSLTETINEPVFKLHQFTLWETIKV	316		
Db				

```

RESULT 13
US-09-328-352-4982
; Sequence 4982, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4982
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4982

```

[illegible]

```

RESULT 14
US-09-629-222A-36
/ Sequence 36 Application US/09629222A
/ Patent No. 659700
/ GENERAL INFORMATION:
/ APPLICANT: Bellacosa, Alfonso
/ TITLE OF INVENTION: Methods for Detection of Transition
/ TITLE OF INVENTION: Single-Nucleotide Polymorphisms
/ FILE REFERENCE: PCCC 96-21
/ CURRENT APPLICATION NUMBER: US/09/629, 222A
/ CURRENT FILING DATE: 2000-07-31
/ PRIOR APPLICATION NUMBER: 09/463, 891
/ PRIOR FILING DATE: 2000-01-28
/ PRIOR APPLICATION NUMBER: PCT/US98/15628
/ PRIOR FILING DATE: 1998-07-28
/ PRIOR APPLICATION NUMBER: 60/053, 936
/ PRIOR FILING DATE: 1997-07-28
/ NUMBER OF SEQ ID NOS: 73
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 36
/ LENGTH: 188
/ TYPE: PRT
/ ORGANISM: E. coli
US-09-629-222A-36

```

	Query Match	22.9%	Score 389.5;	DB 4;	Length 188;	
	Best Local Similarity	44.8%;	Pred No. 3,3e-31;			
	Matches	81;	Conservative	37;	Mismatches	50;
					IndeIs	13; Gaps
						5;
Qy	27 PYRVLTSEVLLOOTRVEQALPYRRRFLEERFPTLKALAASAEVLRWQGAGYVRARHL	86				
	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :					
Db	2 PYKVMISEVWLDOQTQATVIIPYFERPFMARFPVTDLANAPLDEIVLHMTGLGYTARAKRL	61				
Qy	87 HRLARSVEEL-----PPSFAELRGPGLGPRYTAALVAASIAFGERVAAVDGNRYVLSRLF	141				
	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :					
Db	62 HKAAQQVATLHGSKPETHETEEVVALPGVGRSTRGAILSLSLGKHPIIDGNVKRYTLACY	121				
Qy	142 ARRS-EPEKE---LFALAQGLLPF-GVDPGVNQNALBELGTVCCLKRPFGCAQLPAF	195				
	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :					
Db	122 AVSGMGPKKEVKKWSLISGEQVTPAVGVERR--FNQAAMDLDGMTCRSRKPKCSLCPLONG	179				
Qy	196 C 196					
Db	180 C 180					



## RESULT 15

US-09-540-236-2578

; Sequence 2578, Application US/09540236

; Patent No. 6673910

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO MORAXELLA CATAR

; FILE REFERENCE: 2709.2005-001

; CURRENT APPLICATION NUMBER: US/09/540,236

; NUMBER OF SEQ ID NOS: 3840

; SEQ ID NO 2578

; LENGTH: 414

; TYPE: PRT

; ORGANISM: M.catarhalis

US-09-540-236-2578

US-09-540-236-2578

Query Match 20.9%; Score 355; DB 4; Length 414;

Best Local Similarity 28.5%; Pred. No. 3e-27;

Matches 109; Conservative 57; Mismatches 132; Indels 84; Gaps 13;

```

QY 3 AMRKALLAWYRENNAR-PLPMRGKDP-----YRVLSEVLLOOTRVEQALPYVRRFLERP 57
DB 33 SFAKRLITFELHGRHGLPMQIHQPSADIVAVWSEIMLOOTQVTVLKFEFFLARA 92
QY 58 TLKALAAASLEBVLVWOGAGYVRAEHLRLASVE-----LPSFABLRGLPGLG 110
DB 93 TVQELAVADWQEVASFVAGIGYVARRNLHAGAQVADFDTHGRFPETVNEQAVKGVG 152
QY 111 PYTAAVASIAFGERVAADVGNVRVLSRLFA-----RSPKKEKLPALAQGLPEGVD 165
DB 153 RSTAGIVAMGVKKFVIGDGNVKVLAHRAVCGDITGATDKRLMEIATATLTPKEYS- 211
QY 166 GWNQALMELGATVCLPKRPGACPLGAFCRGKEAPGRYPAPRKRAKEE---RLVALV 222
DB 212 GHYAQMMDLGAITICTRTQPKCHLCPVTDCTAVALGVQSOLPVKKKAPPKPHMSTALS 271
QY 223 LIG---RKGVHLERLEGRFQGLYGVPLFPPEELP-----GREAA 258
DB 272 LTHCGTLTLHRONGGIGWDGLMSLPF---MLPLDQNKLDNNTLSDAIFKAWQSDKV 328
QY 259 FGVRSRPLGEV-----RHALT-----RLRVEVRGA---- 285
DB 329 HDLHLSQLEIELPIPTQTLTAYLRHTLTTHVHMLYGMISCLNNSQFNQINQTLTGIDY 388
QY 286 LMEGEGBDPWKRPLPKLMEKVL 307
DB 389 LMT---DTFYNLPLPAAMHKL 407

```

Search completed: January 20, 2005, 06:10:41  
 Job time : 29 secs

**This Page Blank (uspto)**



```

QY      181  LPRKPRGACPLGAFGRKAPGRYPAPRRKAKEERLVALVILGRKGVHLERLEGFQG 240
      181  LPRKPRGACPLGAFGRKAPGRYPAPRRKAKEERLVALVILGRKGVHLERLEGFQG 240
Db      181  LPRKPRGACPLGAFGRKAPGRYPAPRRKAKEERLVALVILGRKGVHLERLEGFQG 240
QY      241  LYGVLFPPEELPGREAAFGVRSRPLGEVRLHTRRLAVRGMALMEGEGDPMKRPLP 300
      241  LYGVLFPPEELPGREAAFGVRSRPLGEVRLHTRRLAVRGMALMEGEGDPMKRPLP 300
Db      241  LYGVLFPPEELPGREAAFGVRSRPLGEVRLHTRRLAVRGMALMEGEGDPMKRPLP 300
QY      301  KLMKVRKALPLLAHAGVPLPDA 325
      301  KLMKVRKALPLLAHAGVPLPDA 325
Db      301  KLMKVRKALPLLAHAGVPLPDA 325

RESULT 2
US-09-925-301-1326
; Sequence 1326, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1326
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (438)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (447)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1326

Query Match      28.4%; Score 483.5; DB 9; Length 486;
Best Local Similarity 36.5%; Pred. No. 4,5e-35;
Matches 135; Conservative 46; Mismatches 106; Indels 83; Gaps 13;

QY      3  AMRKALLAMYRENNARPLPMRG-----EKDPYRVLYSEVLLQOTRVEQALPYRRFLP 54
      3  AMRKALLAMYRENNARPLPMRG-----EKDPYRVLYSEVLLQOTRVEQALPYRRFLP 54
Db      3  AMRKALLAMYRENNARPLPMRG-----EKDPYRVLYSEVLLQOTRVEQALPYRRFLP 54
QY      106  AFRGSLSMWYDQEKRDLPWRRRAEDMDLDRRAVAVSWSEVMQOQVAVIVINYTGMMQ 165
      106  AFRGSLSMWYDQEKRDLPWRRRAEDMDLDRRAVAVSWSEVMQOQVAVIVINYTGMMQ 165
Db      106  AFRGSLSMWYDQEKRDLPWRRRAEDMDLDRRAVAVSWSEVMQOQVAVIVINYTGMMQ 165
QY      55  RPPFLKALAAASLEVEVLWQAGYVRAEHLRLARS-VEED---PPSFAEL--RGLRG 108
      55  RPPFLKALAAASLEVEVLWQAGYVRAEHLRLARS-VEED---PPSFAEL--RGLRG 108
Db      55  RPPFLKALAAASLEVEVLWQAGYVRAEHLRLARS-VEED---PPSFAEL--RGLRG 108
QY      166  KMPFLQDADASLEVEVLWQAGYVRAEHLRLARS-VEED---PPSFAEL--RGLRG 225
      166  KMPFLQDADASLEVEVLWQAGYVRAEHLRLARS-VEED---PPSFAEL--RGLRG 225
Db      166  KMPFLQDADASLEVEVLWQAGYVRAEHLRLARS-VEED---PPSFAEL--RGLRG 225
QY      109  LGPYTAAVASTAFGEVAAVDGNYRVLSRLFA-----RSPKEXELPLAQLPBGV 163
      109  LGPYTAAVASTAFGEVAAVDGNYRVLSRLFA-----RSPKEXELPLAQLPBGV 163
Db      109  LGPYTAAVASTAFGEVAAVDGNYRVLSRLFA-----RSPKEXELPLAQLPBGV 163
QY      226  VERTYTAGAISTAFGAGTGVVDGNYRVLSRLFA-----RSPKEXELPLAQLPBGV 284
      226  VERTYTAGAISTAFGAGTGVVDGNYRVLSRLFA-----RSPKEXELPLAQLPBGV 284
Db      226  VERTYTAGAISTAFGAGTGVVDGNYRVLSRLFA-----RSPKEXELPLAQLPBGV 284
QY      164  DRGVNVALMELGATVCLPKRPRGACPLGAFGRKAKEERLVALVILGRKGVHLERLEGFQG 344
      164  DRGVNVALMELGATVCLPKRPRGACPLGAFGRKAKEERLVALVILGRKGVHLERLEGFQG 344
Db      164  DRGVNVALMELGATVCLPKRPRGACPLGAFGRKAKEERLVALVILGRKGVHLERLEGFQG 344
QY      285  RPDDEFQOAMELGATVCLPKRPRGACPLGAFGRKAKEERLVALVILGRKGVHLERLEGFQG 344
      285  RPDDEFQOAMELGATVCLPKRPRGACPLGAFGRKAKEERLVALVILGRKGVHLERLEGFQG 344
Db      285  RPDDEFQOAMELGATVCLPKRPRGACPLGAFGRKAKEERLVALVILGRKGVHLERLEGFQG 344
QY      202  P-----GRYPARRKRAK-----EELVALYL-----LGRKGVHLER 233
      202  P-----GRYPARRKRAK-----EELVALYL-----LGRKGVHLER 233
Db      202  P-----GRYPARRKRAK-----EELVALYL-----LGRKGVHLER 233
QY      345  PNTGQCHLCPREEPWDTGLGVNFPKRSRKRPRESSATVLEQPALGAILVORP 404
      345  PNTGQCHLCPREEPWDTGLGVNFPKRSRKRPRESSATVLEQPALGAILVORP 404
Db      345  PNTGQCHLCPREEPWDTGLGVNFPKRSRKRPRESSATVLEQPALGAILVORP 404
QY      234  LSGRFGGLGVN--LPPPEELPGREAAF-----GVRSRPLGEVRLHTRRLAV 280
      234  LSGRFGGLGVN--LPPPEELPGREAAF-----GVRSRPLGEVRLHTRRLAV 280
Db      234  LSGRFGGLGVN--LPPPEELPGREAAF-----GVRSRPLGEVRLHTRRLAV 280
QY      405  NSGLLAGLWEPFVTEPSEOLQRKALLQELQRXAGPLPATXIRHLGCVHTFSHIKLTY 464
      405  NSGLLAGLWEPFVTEPSEOLQRKALLQELQRXAGPLPATXIRHLGCVHTFSHIKLTY 464
Db      405  NSGLLAGLWEPFVTEPSEOLQRKALLQELQRXAGPLPATXIRHLGCVHTFSHIKLTY 464

```

```

QY      281  EYRGALMEGE 290
      281  EYRGALMEGE 290
Db      465  QYVGLALBEGQ 474
      465  QYVGLALBEGQ 474

RESULT 3
US-10-289-762-421
; Sequence 421, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 421
; LENGTH: 375
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-289-762-421

Query Match      28.0%; Score 477; DB 15; Length 375;
Best Local Similarity 38.6%; Pred. No. 1,3e-34;
Matches 117; Conservative 44; Mismatches 112; Indels 30; Gaps 7;

QY      6  KALLAMYRENNARPLPMRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAA 65
      6  KALLAMYRENNARPLPMRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAA 65
Db      6  KALLAMYRENNARPLPMRGKDPYRVLYSEVLLQOTRVEQALPYRRFLERFPTLKALAAA 65
QY      22  EALKKMFERNKRSPLPMRNDPTPYVSWSEVMQOQVAVIVINYTGMMQ 81
      22  EALKKMFERNKRSPLPMRNDPTPYVSWSEVMQOQVAVIVINYTGMMQ 81
Db      22  EALKKMFERNKRSPLPMRNDPTPYVSWSEVMQOQVAVIVINYTGMMQ 81
QY      66  SLEEVLYRWQAGYVRAEHLRLARSVEE---LPPSFAELRGLRGAGPYTAAVAST 120
      66  SLEEVLYRWQAGYVRAEHLRLARSVEE---LPPSFAELRGLRGAGPYTAAVAST 120
Db      66  SLEEVLYRWQAGYVRAEHLRLARSVEE---LPPSFAELRGLRGAGPYTAAVAST 120
QY      82  KEEDVYIKLMESGLGYVSRARHLLEGARWMEBFHGKIDDAISLAQIRGVSPYVHAILAR 141
      82  KEEDVYIKLMESGLGYVSRARHLLEGARWMEBFHGKIDDAISLAQIRGVSPYVHAILAR 141
Db      82  KEEDVYIKLMESGLGYVSRARHLLEGARWMEBFHGKIDDAISLAQIRGVSPYVHAILAR 141
QY      121  AFGERYAAVDGNYRVLSRLFAEESPEKE-----LPALAGLLPBGVGNVALMEL 175
      121  AFGERYAAVDGNYRVLSRLFAEESPEKE-----LPALAGLLPBGVGNVALMEL 175
Db      121  AFGERYAAVDGNYRVLSRLFAEESPEKE-----LPALAGLLPBGVGNVALMEL 175
QY      142  AFRRAAAVDGNYRVLSRLFAEESPEKE-----LPALAGLLPBGVGNVALMEL 200
      142  AFRRAAAVDGNYRVLSRLFAEESPEKE-----LPALAGLLPBGVGNVALMEL 200
Db      142  AFRRAAAVDGNYRVLSRLFAEESPEKE-----LPALAGLLPBGVGNVALMEL 200
QY      176  GATVCLPKRPRGACPLGAFGRKAPGRYPAPRRKAKEERLVALVILGRKGVHLERLEGFQG 230
      176  GATVCLPKRPRGACPLGAFGRKAPGRYPAPRRKAKEERLVALVILGRKGVHLERLEGFQG 230
Db      176  GATVCLPKRPRGACPLGAFGRKAPGRYPAPRRKAKEERLVALVILGRKGVHLERLEGFQG 230
QY      201  GACIC-KCTPQCHRCVRCAGAMRENKQFVLPVHARKVIFLRLVALVLDGSLVE 259
      201  GACIC-KCTPQCHRCVRCAGAMRENKQFVLPVHARKVIFLRLVALVLDGSLVE 259
Db      201  GACIC-KCTPQCHRCVRCAGAMRENKQFVLPVHARKVIFLRLVALVLDGSLVE 259
QY      231  LERLEGRFGGLGVN--LPPPEELPGREAAFGVRSR-----PLGEVRLHTRRLAV 277
      231  LERLEGRFGGLGVN--LPPPEELPGREAAFGVRSR-----PLGEVRLHTRRLAV 277
Db      231  LERLEGRFGGLGVN--LPPPEELPGREAAFGVRSR-----PLGEVRLHTRRLAV 277
QY      260  KRRPKEMMAGLYEFYIYEVPESEGQDLEGFTKQNELSLESPLELGLVKEQRHAFTHVK 319
      260  KRRPKEMMAGLYEFYIYEVPESEGQDLEGFTKQNELSLESPLELGLVKEQRHAFTHVK 319
Db      260  KRRPKEMMAGLYEFYIYEVPESEGQDLEGFTKQNELSLESPLELGLVKEQRHAFTHVK 319
QY      278  LRV 280
      278  LRV 280
Db      278  LRV 280
QY      320  VHL 322
      320  VHL 322
Db      320  VHL 322

RESULT 4
US-10-425-115-27736
; Sequence 27736, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 27736
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Zea mays

```

FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_184509C.1.pep  
US-10-425-115-277336

Query Match 26.3%; Score 448; DB 17; Length 469;  
Best Local Similarity 36.4%; Pred. No. 7.2e-32;  
Matches 123; Conservative 44; Mismatches 113; Indels 58; Gaps 10;

QY 3 AMRKALLARYRENARPLPWR-----GKDPYRLVSEVLLQOTRVEQALPYRRFLERP 57  
DB 57 ALRAQQLRWYDARRDLPMRCVSGSEEBRAYAVWVSEVMLQOTRVPVVAAYYERMMARWP 116  
QY 58 TLKALAAASLEEVLRVWOGAGYRRAEHLRLARSVE---LPPSFALRGFLGCLPYTA 114  
DB 117 TVRSLLAAATQOEVENEMAGLYRRARFLLEGAKQIIEKLFECTALALREVGIGIDYTA 176  
QY 115 AAVASIAFGERVAADVGNVRLSLFA--RESPKEKELFALAGLLPEGVD---PGVMNQ 170  
DB 177 GAIASIAFNBVVVDGNVIRVLSRLYTADNPKESSTVVRPMDLVGQWVDPLRPGDFNQ 236  
QY 171 ALMELGATVCLPKRPGCAGPLGAF-----RGKAPGYRPA---KRRAEERLVAL 221  
DB 237 AMMELGATLCSKTKPGCSQCPVSHCOALALSREKSSVQVTDPPRVVPKAKPRSDPAAYC 296  
QY 222 VLLGRGV-----HLERL-----EGRFQGLYGVLPFPE---ELPGRBAFG- 260  
DB 297 VVOIAGLEBEAADPKGNHLLILKRPBEGLAGLMEFPVLVVDGKTDLNRRKAMD 356  
QY 261 -----VRSRPLGEVRLHRLRLRYEV 282  
DB 357 YLSKLISIDMVRKPDVILREDVEHVFHSIRLTMHV 394

RESULT 5  
US-10-425-114-68280  
Sequence 68280, Application US/10425114  
Publication No. US20040034888A1

GENERAL INFORMATION:  
APPLICANT: Liu, Jindong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaka, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(5313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 68280  
LENGTH: 482  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: UC-ZMFLB7326SD07\_FLI.pep  
US-10-425-114-68280

Query Match 26.3%; Score 448; DB 15; Length 482;  
Best Local Similarity 36.4%; Pred. No. 7.5e-32;  
Matches 123; Conservative 44; Mismatches 113; Indels 58; Gaps 10;

QY 3 AMRKALLARYRENARPLPWR-----GKDPYRLVSEVLLQOTRVEQALPYRRFLERP 57  
DB 70 ALRAQQLRWYDARRDLPMRCVSGSEEBRAYAVWVSEVMLQOTRVPVVAAYYERMMARWP 129  
QY 58 TLKALAAASLEEVLRVWOGAGYRRAEHLRLARSVE---LPPSFALRGFLGCLPYTA 114  
DB 130 TVRSLLAAATQOEVENEMAGLYRRARFLLEGAKQIIEKLFECTALALREVGIGIDYTA 189  
QY 115 AAVASIAFGERVAADVGNVRLSLFA--RESPKEKELFALAGLLPEGVD---PGVMNQ 170  
DB 190 GAIASIAFNBVVVDGNVIRVLSRLYTADNPKESSTVVRPMDLVGQWVDPLRPGDFNQ 249

QY 171 ALMELGATVCLPKRPGCAGPLGAF-----RGKAPGYRPA---KRRAEERLVAL 221  
DB 250 AMMELGATLCSKTKPGCSQCPVSHCOALALSREKSSVQVTDPPRVVPKAKPRSDPAAYC 309  
QY 222 VLLGRGV-----HLERL-----EGRFQGLYGVLPFPE---ELPGRBAFG- 260  
DB 310 VVOIAGLEBEAADPKGNHLLILKRPBEGLAGLMEFPVLVVDGKTDLNRRKAMD 369  
QY 261 -----VRSRPLGEVRLHRLRLRYEV 282  
DB 370 YLSKLISIDMVRKPDVILREDVEHVFHSIRLTMHV 407

RESULT 6  
US-10-437-963-138271  
Sequence 138271, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(5322)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 138271  
LENGTH: 474  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_39677C.1.pep  
US-10-437-963-138271

Query Match 26.0%; Score 442; DB 16; Length 474;  
Best Local Similarity 35.0%; Pred. No. 2.6e-31;  
Matches 122; Conservative 47; Mismatches 110; Indels 70; Gaps 12;

QY 3 AMRKALLARYRENARPLPWRGKDP-----YRLVSEVLLQOTRVEQALPY 48  
DB 56 AVRABELLRWYDARRDLPMRBAEPAGSGRGEQRAYAVWVSEVMLQOTRVPVVDY 115  
QY 49 YRRFLERPTLKALAAASLEEVLRVWOGAGYRRAEHLRLARSV---BELPSPFALRG 105  
DB 116 YSRMARWPTVDSLAATOEEVENEMAGLYRRARFLLEGAKQIIEKSEFPCTASTLRE 175  
QY 106 LGLGPTYTAAAVASIAFGERVAADVGNVRLSLFA--RESPKEKELFALAGLLPEGVD 164  
DB 176 VRSIGDYTAGALASIAFNBVVVDGNVIRVLSRLYTADNPKESSTVVRPMDLVGQWVDPLRPGDFNQ 236  
QY 165 -----PGVMNQALMELGATVCLPKRPGCAGPLGAFRGKAPGR-----YP--APRKR 211  
DB 236 PSRPGDFNQMAMELGATLCSKTKRPGCSQCPVSHCOALALSQNASVKTDFPRVVPKAK 295  
QY 212 RAKERLVALVLLGRKGVHLERLEGRFQGLYGVLPFPEBELPGRBAFGVRSRPLGEVRH 271  
DB 296 PRSDPAAYCVVOIS-----QGRGEGI-----AEAEKDNLLFLIKRP----- 332  
QY 272 ALTHRLRLREVGALME-----GEGE--DPWRKPLPKLMEKVLKALPL 313  
DB 333 -----BEGLAGLMEFPVSVLVBEGKTDTLNR--RKENDKYLKQULSI 372

RESULT 7  
US-10-629-951-36  
Sequence 36, Application US/10629951

Query Match	22.9%;	Score 389.5;	DB 15;	Length 188;
Best Local Similarity	44.8%;	Pred. No. 4.5e-27;		
Matches	81;	Conservative	37;	Mismatches 50;
				Indels 13;
				Gaps 5

RESULT 8  
US-10-156-761-12241

Query Match	Score	DB	Length
22.2%	377	14	313

RESULTS  
US-09-738-626-6433  
; Sequence 6433, Application US/09738626  
; Publication No. US20020197605A1

Query Match	21.5%;	Score 365;	DB 9;	Length 293;
Best Local Similarity	34.6%;	Pred. No. 1.3e-24;		
Matches	99;	Conservative 47;	Mismatches 120;	Indels 20;
			Gaps	7

122 ARAVAAFHGQRPVDTNVRVYQRAVAGRYLAGPAKKQELIDVSLLENITHAF-EFSA 10



Db 61 QLKELARVINDYGVRRNRKALIDLPVGKTYCAAVMCLAFGKAAVDANFVRVYNR 120  
QY 140 LFARE----SPKRELFAALQGLPEGVDPGVNQAALMELGATVCLPRPRCGACPLGAF 195  
Db 121 YFGGSYENLNNHMLMELATVLPVGGKCRD-FILGLMDFSAIICAPRKCCEKCGMSKL 179  
QY 196 C 196  
Db 180 C 180

RESULT 13  
US-10-335-977-7283  
Sequence 7283, Application US/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 7283:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 230 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...230  
SEQUENCE DESCRIPTION: SEQ ID NO: 7283:  
US-10-335-977-7283

Query Match 13.8%; Score 234; DB 15; Length 230;  
Best Local Similarity 30.6%; Pred. No. 7,7e-13;  
Matches 66; Conservative 33; Mismatches 54; Indels 64; Gaps 9;

QY 95 ELPPSFARLRLGQLPGYTAAVAASIAFGERVAADVGNVRVLSRLFARESPEKELFAL 154  
Db 7 QLPNDYQSLTKLPGIGAYTANAILCFGRFKSAQVDANVRVLRLLRF----- 53  
QY 155 AAGLLPEGVDPGV-----WQAALMELGATVCLPRPRCGACPLGA 194  
Db 54 -----GLDPNTHAKDLQIKANDPLNLNFSFNHNOALIDLGLALISPK-PKCAICPFNP 105

QY 195 FCRGKEAPGRYPAPRRR-AKEERLVALVLLGRKGVHLERLEGRFOGLY-GVPLFP----- 248  
Db 106 YCLGKHLNHLHTLKKQEIIGERYIGVYQNNQ-IALEKIE---QKVLGCHNHFNNLKE 161  
QY 249 --PEELGREANFVRSRPLGEVRAHLLTHRRRLREV 282  
Db 162 NLEFKLPF-----LGIKHSHTYKKNLNL 186

RESULT 14  
US-10-335-977-7282  
Sequence 7282, Application US/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 7282:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 187 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...187  
SEQUENCE DESCRIPTION: SEQ ID NO: 7282:  
US-10-335-977-7282

Query Match 13.4%; Score 228.5; DB 15; Length 187;  
Best Local Similarity 33.5%; Pred. No. 1.9e-12;  
Matches 59; Conservative 25; Mismatches 45; Indels 47; Gaps 7;

QY 95 ELPPSFARLRLGQLPGYTAAVAASIAFGERVAADVGNVRVLSRLFARESPEKELFAL 154  
Db 16 QLPNDYQSLTKLPGIGAYTANAILCFGRFKSAQVDANVRVLRLLRF----- 62  
QY 155 AAGLLPEGVDPGV-----WQAALMELGATVCLPRPRCGACPLGA 194  
Db 63 -----GLDPNTHAKDLQIKANDPLNLNFSFNHNOALIDLGLALISPK-PKCAICPFNP 114  
QY 195 FCRGKEAPGRYPAPRRR-AKEERLVALVLLGRKGVHLERLEGRFOGLY-GVPLFP 248



DB 115 YCIGKHLRHTLKKQELIQEBRYLGVVIONNQ-IALEKIE---QKLYLGMHFP 166

## RESULT 15

US-10-335-977-7279

; Sequence 7279, Application US/10335977  
; Publication No. US20040052799A1

## GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR

DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSER: LAHIVE &amp; COCKFIELD

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: Windows NT 4.0

SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977

FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragoras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 7279:

SEQUENCE CHARACTERISTICS:

LENGTH: 98 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...98

SEQUENCE DESCRIPTION: SEQ ID NO: 7279:

US-10-335-977-7279

Query Match 11.8%; Score 200.5; DB 15; Length 98;

Best Local Similarity 48.4%; Pred. No.2.8e-10;

Matches 45; Conservative 14; Mismatches 29; Indels 5; Gaps 3;

QY 1 MEAMRRALLAWYRENR-PLPWR--GKDPYRVLVSEVLLQOTRVEQALP-YRRFLER 55

DB 1 LETLNHALIKWYEFGRKDLPPFNKGINAPYEVYISEVWSQOTQISTVERFYPPFLKA 60

QY 56 PPTLKALAAASLEVRVWQAGAGYRRAEHLR 88

DB 61 PPTLKDLANAPLEEVILLWRGLGYSRANKLKK 93

Search completed: January 20, 2005, 06:16:04

Job time : 82 secs

**This Page Blank (uspto)**